

**KÍSÉRLETTERVEZÉS ÉS ÉRTÉKELÉS A MIKROBIOLÓGIAI  
GYAKORLATBAN**

**DR. REICHART OLIVÉR**

**2005.**

**Budapest**

**Lektorálta: Zukál Endre**

## Tartalom

BEVEZETÉS	3
1. VALÓSZÍNŰÉGSZÁMÍTÁSI ALAPOK	5
1.1. Kombinatorikai alapösszefüggések	5
1.2. A valószínűség fogalma	6
1.2.1. Események közötti műveletek és összefüggések	6
1.2.2. Eseményalgebrai műveletek	7
1.2.3. A valószínűségre vonatkozó néhány tétel	7
1.2.4. Feltételes valószínűség, független események	8
1.2.5. Teljes valószínűség tétele, Bayes tétel	8
1.3. Valószínűségi változó	9
1.4. Valószínűségi változó eloszlásfüggvénye és sűrűségfüggvénye	9
1.5. Valószínűségi változó főbb jellemzői	10
1.5.1. Várható érték ( $M$ )	10
1.5.2. Szórás ( $\sigma$ )	11
1.6. Standardizált valószínűségi változó	12
1.7. Csebisev egyenlőtlenség	12
1.8. A valószínűségi eloszlás általános jellemzői	13
1.9. A valószínűségi eloszlás paramétereinek becslése	14
2. VALÓSZÍNŰÉGI ELOSZLÁSOK	15
2.1. Diszkrét eloszlások	15
2.1.1. Egyenletes eloszlás	15
2.1.2. Binomiális (Bernoulli) eloszlás	16
2.1.3. Hipergeometrikus eloszlás	19
2.1.4. Poisson eloszlás	19
2.2. Folytonos eloszlások	21
2.2.1. Normális eloszlás	21
2.2.2. Student eloszlás	23
2.2.3. $\chi^2$ eloszlás	24
2.2.4. F eloszlás	25
2.2.5. Exponenciális eloszlás	26
3. STATISZTIKAI BECSLÉSEK	27
3.1. A minta statisztikai jellemzői	30
3.1.1. Tapasztalati eloszlás meghatározása	30
3.1.2. Várható érték becslésére szolgáló jellemzők	38
3.1.3. Szórás becslésére szolgáló jellemzők	40
3.2. Az alapsokaság paramétereinek becslése	42
3.2.1. Normális eloszlású alapsokaság paramétereinek becslése	42
3.2.2. Binomiális eloszlás paramétereinek becslése	42
3.2.3. Poisson eloszlás $\lambda$ paraméterének becslése	42
3.3. Konfidencia-intervallum számítások	42
3.3.1. Normál eloszlással kapcsolatos konfidencia-intervallum számítások	43

3.3.1.1. Várható érték konfidencia-intervalluma ismert szórás esetén	43
3.3.1.2. Szórás és várható érték konfidencia-intervalluma ismeretlen szórás esetén	44
3.3.2. Binomiális eloszlással kapcsolatos konfidencia-intervallum számítások	47
3.3.3. Poisson eloszlással kapcsolatos konfidencia-intervallum számítások	48
<b>4. HIPOTÉZIS VIZSGÁLATOK, STATISZTIKAI PRÓBÁK</b>	<b>50</b>
4.1. Egymintás statisztikai próbák	51
4.1.1. Szórás összehasonlítása elméleti értékkel	51
4.1.2. Poisson eloszlású sokaság várható értékének statisztikai próbái	52
4.1.3. Normál eloszlású sokaság várható értékének statisztikai próbái	53
4.1.3.1. Ismert szórású normál eloszlású sokaság statisztikai próbája	53
4.1.3.2. Ismeretlen szórású normál eloszlású sokaság statisztikai próbája	53
4.2. Kétmintás statisztikai próbák	54
4.2.1. Két számított szórás összehasonlítása	54
4.2.2. Poisson eloszlású sokaságok várható értékének összehasonlítása	55
4.2.3. Normál eloszlású sokaságok várható értékének összehasonlítása	56
4.2.3.1. Azonos szórású normál eloszlású sokaságok összehasonlítása	56
4.2.3.2. Eltérő szórású normál eloszlású sokaságok összehasonlítása	56
4.2.3.3. Párosított adatok eltérésére vonatkozó statisztikai próba	58
4.3. Többszörös statisztikai próbák	59
4.3.1. Poisson eloszlású gyakoriságok összehasonlítása	60
4.3.2. Szórások összehasonlítása	61
4.3.3. Várható értékek összehasonlítása variancia analízissel	62
4.3.3.1. Egyszempontos variancia analízis	64
4.3.3.2. Kétszempontos variancia analízis	66
4.3.3.2.1. Kétszempontos variancia-analízis ismétlések nélküli elrendezéssel	66
4.3.3.2.2. Kétszempontos variancia-analízis kezelésen belüli párhuzamosokkal	68
<b>5. LINEÁRIS REGRESSZIÓ</b>	<b>71</b>
<b>6. ÉLŐSEJTSZÁM MEGHATÁROZÁSI MÓDSZEREK HIBÁJA</b>	<b>76</b>
6.1. Határhígításos (MPN) módszer	76
6.2. Lemezöntéses és felületi szélesztéses módszer	80
6.3. Határhígításos és lemezöntéses módszerek összehasonlítása	85
6.4. Módszer ismételhőségének és reprodukálhatóságának meghatározása	86
<b>MATEMATIKAI-STATISZTIKAI TÁBLÁZATOK</b>	<b>95</b>
<b>IRODALOM</b>	<b>111</b>

## BEVEZETÉS

A laboratóriumi gyakorlatban a kísérlettervezés és -értékelés célja a valóság "kihámozása" a mérési (megfigyelési) adatokból, eszköztára a matematikai-statisztika, amely a valószínűségszámításból, annak speciális területeként alakult ki.

A valószínűségszámítás *véletlen tömegjelenségekkel* foglalkozik. Véletlen tömegjelenségek alatt olyan jelenségeket, eseményeket értünk, amelyek azonos körülmények között nagyon nagy számban fordulnak elő (akár egyidejűleg, vagy időben egymásután), illetőleg elvileg tetszőlegesen sokszor megismételhetők. Az első csoportba sorolhatók be pl. a tömeggyártás során előállított termékek (csomagolt élelmiszerek, stb.) raktározott tételeinek egyedei, vagy a gyártó gépről lejövő egyes darabjai. A második csoport jellegzetes képviselői a laboratóriumi mérési eredmények.

A tömegjelenség véletlen volta azt jelenti, hogy a megfigyelés, mérés (a továbbiakban kísérlet) eredményét nagyon sok tényező befolyásolja. Ezen tényezők esetleg ismertek, vagy ismeretlenek, s legalább egy részük időben változik. A kísérlet eredménye a befolyásoló tényezők együttes hatásaként, véletlenszerűen alakul ki. Amennyiben az eredményt befolyásoló tényezők időben változatlanok, nem okoznak véletlenszerű ingadozást. A kísérlettervezés és -értékelés célja nagyon sok esetben éppen ezeknek az állandó (rendszeres) hatásoknak a véletlen hatásoktól való elválasztása, illetve kimutatása.

A kísérleti eredmények matematikai-statisztikai kiértékelése lehetővé teszi számunkra, hogy mérési eredményeink alapján meghatározzuk a véletlenre visszavezethető ingadozások mértékét, s ennek figyelembevételével hozzuk meg döntéseinket. Ezek a döntéseink elméletileg sohasem 100%-os biztonságúak, mindig magukban rejtik a tévedés bizonyos valószínűségét. Hogy ez a tévedési valószínűség milyen mértékű, azt a kísérleti elrendezés és a mért jellemző véletlen ingadozása határozza meg.

A kísérleti eredmények kiértékelésében alkalmazott számítási módszerek általában függetlenek a kísérletek fizikai, kémiai, biológiai jellegétől, a mért jellemzőtől, ezért a matematikai-statisztika teljesen önállóan, mint a matematika egyik résztudományága kezelhető. A matematikai-statisztikai eljárások azonban valószínűség-számítási alapjaik miatt csak olyan esetben vezetnek korrekt következtetésekre, amikor az alapadatként felhasznált jellemző megfelel bizonyos matematikai feltételeknek. Ha ezek a feltételek nem teljesülnek, a kiértékelés eredménye, s az ebből levonható következtetések rendkívül félrevezetőek lehetnek.

A tudományos kutatásban, vagy a rutin laboratóriumi munkában a kísérletek célja mindig valamilyen kérdésre adandó válasz meghatározása, vagy valamely összefüggés megállapítása (mekkora egy fermentlé enzimaktivitása, cukortartalma, van-e különbség két táptalajon meghatározott mikrobaszám között, megfelel-e a vizsgált termék az érvényben lévő minőségi előírásoknak, hogyan függ egy mikroba hőpusztulási ideje a hőmérséklettől, stb.). Költséges, időigényes kísérleti módszerek esetén egyáltalán nem mindegy, hogy hány méréssel, ill. mennyi idő alatt és milyen *megbízhatósággal* adjuk meg a választ a feltett kérdésekre, ezért lényeges, hogy milyen *kísérleti elrendezést* alkalmazunk. A jó kísérleti elrendezés igen nagy mértékben megnöveli a kiértékelés hatékonyságát, illetve jelentősen csökkenti az adott biztonságú döntéshez szükséges mérések számát.

A kísérleti eredmények feldolgozásában és kiértékelésében ma már egyre elterjedtebben alkalmazzák a számítógépes eljárásokat, amelyek igen nagy mértékben megkönnyítik a számításokat. Vitathatatlan előnyeik ellenére a statisztikai szubrutinok gépies alkalmazása két alapvető veszélyforrást is rejt magában:

A kísérleti elrendezés kiválasztásában nagyon sokszor nem annak célszerűsége, hanem a rendelkezésre álló matematikai-statisztikai szubrutin jellege a döntő. Ez a kiértékelés hatékonyságának (a döntés biztonságának) csökkenéséhez, vagy a kísérletek számának felesleges növeléséhez vezethet.

A kiértékelésbe bevont alapadatoknál nagyon sok esetben elmarad az alkalmazott számítási módszer által megkívánt feltételek teljesülésének vizsgálata, s ennek eredményeként a levont következtetések tévesek lehetnek.

A fenti két hibaforrás bármely (nem csak a számítógépes) matematikai-statisztikai kiértékelésnél előfordulhat, ezért a továbbiakban mindvégig súlyt fektetünk a számítási eljárások alapfeltételeinek ismertetésére, valamint az esetenkénti leghatékonyabb kísérleti elrendezés bemutatására.

A matematikai-statisztikai módszerek, - ahogy azt már előzőleg említettük, - általános érvénnyel használhatók az általános laboratóriumi gyakorlatban, mikrobiológiai adatokra vonatkozóan azonban néhány alapvető szempontra szükséges felhívni a figyelmet.

A fizikai, kémiai jellemzők (tömeg, koncentráció, vezetőképesség, stb.) ingadozása a külső körülmények standardizálásával nagy mértékben csökkenthető. Ezzel szemben a mikrobiológiai jellegű méréseknél egy plusz "bizonytalansági" tényező mindig megmarad, ami a mérendő paraméter *biológiai* jellegéből fakad. Ebből eredően a mikrobiológiai mérések véletlen ingadozása relatíve jelentősen nagyobb a fizikai-kémiai jellegű paraméterek ingadozásánál.

Mikrobaszám meghatározása esetében a mért tartomány rendkívül nagy lehet (esetenként  $10^0$ - $10^9$  sejt/ml), összehasonlítva pl. a kémiai paraméterek 1-2 nagyságrend szélességű változásával.

A fizikai és kémiai jellegű mérések eredményei általában teljesítik a matematikai-statisztikai módszerek alkalmazásának alapadatokra vonatkozó feltételeit. Ugyanez mikrobiológiai eredmények esetében a legritkábban fordul elő. A mikrobaszámokat feldolgozás előtt matematikailag transzformálni kell, s a számításokat ezekkel az (általában log-transzformált) adatokkal kell elvégezni.

Az előbbieken ismertetett tulajdonságok csupán a mikrobiológiai eredmények kissé "különleges" voltára utalnak, de mint már említettük, megfelelő figyelembevételük után az általános érvényű matematikai-statisztikai eljárások teljes körűen alkalmazhatók.

Bevezetésünkben végül egy igen fontos szempontra kívánjuk felhívni a figyelmet. A matematikai-statisztikai módszerek alkalmazása a döntéshozásban igen hatékony segédeszköz, de sohasem válhat öncélúvá, nem helyettesítheti a szakmai döntéseket. A matematikai-statisztikai döntések arról adnak felvilágosítást, hogy véletlen tömegjelenséggként kezelt kísérleti eredmények valószínűségszámítási megfontolások figyelembevételével hogyan értelmezhetők. Az eredmények alapján hozott szakmai döntéseket a matematikai értékelés nem helyettesíti, csupán alátámasztja.

# 1. VALÓSZÍNŰÉGSZÁMÍTÁSI ALAPOK

## 1.1. KOMBINATORIKAI ALAPÖSSZEFÜGGÉSEK

A kombinatorika alapösszefüggései, melyeket a valószínűségyszámításban igen gyakran felhasználunk, különböző, esetleg részben azonos elemek elrendezhetőségének szabályaival foglalkozik (pl. hányféleképpen választható ki  $n$  elemű tételből  $k$  db. minta, stb.)

Fejezetünkben csupán az alapösszefüggéseket ismertetjük, azok levezetése az irodalomjegyzékben megadott könyvekben részletesen megtalálható.

### Permutáció

*Ismétlés nélküli permutáció*

$n$  különböző elem különböző sorrendjének a számát ( $P_n$ ) adja meg:

$$P_n = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n = n! \quad (\text{olvasd: } n \text{ faktoriális}) \quad (1.1./1.)$$

*Ismétléses permutáció*

Ha az  $n$  elemen belül  $k_1, k_2, \dots, k_l$  darab egyező van, az  $n$  elem ismétléses permutációinak száma:

$$P_n^{k_1, k_2, \dots, k_l} = \frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_l!} \quad (1.1./2.)$$

### Variáció

*Ismétlés nélküli variáció*

$n$  különböző elemből  $k$  db. különbözőt kiválasztunk ( $k \leq n$ ) és minden lehetséges sorrendbe állítjuk, akkor az így keletkező variációk száma:

$$V_{n,k} = n(n-1)(n-2)\dots(n-k+1) \quad (1.1./3.)$$

*Ismétléses variáció*

Ha a kiválasztásnál ismétlődést is megengedünk, az ismétléses variációk száma:

$$V_{n,k}^{ism} = n^k \quad (1.1./4.)$$

### Kombináció

*Ismétlés nélküli kombináció*

Ha  $n$  különböző elemből  $k$  db-ot kiválasztunk ( $k \leq n$ ), de a kiválasztott elemeket nem rakjuk különböző sorrendbe, a keletkező kombinációk száma:

$$C_{n,k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot k} = \binom{n}{k} \quad (1.1./5.)$$

Megegyezés szerint  $\binom{n}{0} = 1$ .  $(1.1./6.)$ .  $\binom{n}{1} = n$   $(1.1./7.)$

### Ismétléses kombináció

Ha  $n$  elemből  $k$  db-ot kiválasztunk úgy, hogy az egyes elemeket többször is kiválaszthatjuk, az így kapható ismétléses kombinációk száma:

$$C_{n,k}^{ism} = \binom{n+k-1}{k} \quad (1.1./8.)$$

## 1.2. A VALÓSZÍNŰSÉG FOGALMA

A valószínűségszámítás véletlen tömegjelenségekkel foglalkozik. A véletlen jelenségek megfigyelésére szolgál a *kísérlet*, melynek lehetséges kimeneteleit (eredményeit) *eseményeknek* nevezzük. A továbbiakban csak olyan kísérletekkel foglalkozunk, melyek azonos körülmények között elvileg végtelen sokszor megismételhetők.

A kísérletek eredményeként kapott események lehetnek minőségileg különbözőek (pl. egy kártyacsomag különböző lapjai, egy termék megfelelő vagy selejtes volta), vagy azonos minőségű, de eltérő mennyiségek (pl. a kockadobás eredménye, egy sűrítmény cukortartalma).

Ha egy  $n$  alkalommal elvégzett kísérletből a kiszemelt  $A$  esemény  $k$ -szor következik be, akkor a  $k$  szám az  $A$  esemény *gyakorisága*,  $k/n$  pedig a *relatív gyakorisága*.

A kísérlet véletlen jellege abban nyilvánul meg, hogy a kísérletsorozatot azonos körülmények között többször megismételve, az  $A$  esemény relatív gyakoriságára eltérő értékeket kapunk. Ezek a relatív gyakoriságok egy meghatározott számérték körül ingadoznak, mégpedig annál kisebb mértékben, minél nagyobb az  $n$  értéke. Az  $A$  esemény relatív gyakoriságának ezt a várható számértékét az  **$A$  esemény valószínűségének  $P(A)$**  nevezzük.

Matematikai megfogalmazással **a relatív gyakoriság határértéke a valószínűség**.

A relatív gyakoriság és a valószínűség 0 és 1 közé eső számok. A biztos esemény valószínűsége 1, a lehetetlené 0. A kísérlet minden egyes kimeneteléhez kiszámítható a relatív gyakoriság, amely az adott esemény valószínűségének becslésére szolgál.

### 1.2.1. Események közötti műveletek és összefüggések

Egy kísérlet lehetséges eredményeit elemi eseményeknek nevezzük. Az elemi események halmaza az *I eseménytér*. Az esemény az eseménytér egy részhalmaza. Ilyen értelmezésben a kísérleti eredményekre is alkalmazhatóak az eseményalgebra összefüggései.

Az események jelölésére általában nagybetűket használunk (pl.  $A$  esemény,  $B$ ,  $B_1$ ,  $B_2$  események, stb.).

- Az összes eseményt tartalmazó halmaznak az **I** biztos esemény felel meg.
- Egy **A** esemény **ellentéte** az  $\bar{A}$  esemény, amely kizárólag akkor következik be, ha  $A$  nem következik be
- A biztos esemény ellentéte a  $\emptyset$  lehetetlen esemény



Az  $A_1, A_2, A_3 \dots, A_n$  események teljes eseményrendszert alkotnak, ha egyikük biztosan bekövetkezik, és ha egymást páronként kizárják. Az eseménytér összes lehetséges elemi eseménye teljes eseményrendszert alkot.

### 1.2.2. Eseményalgebrai műveletek

**Összeadás:** Az  $A+B$  esemény bekövetkezése azt jelenti, hogy vagy  $A$ , vagy  $B$  esemény bekövetkezik, vagy mindkettő.

**Szorzás:** Az  $A \cdot B$  esemény bekövetkezik, ha  $A$  is és  $B$  is bekövetkezik

**Kivonás:** Az  $A-B$  esemény akkor következik be, ha  $A$  bekövetkezik, de  $B$  nem, azaz:  
 $(A - B = A \cdot \bar{B})$ .

### Eseményalgebrai műveleti azonosságok

- a./  $A+B = B+A; \quad A \cdot B = B \cdot A$
- b./  $A+(B+C) = (A+B)+C; \quad A \cdot (B \cdot C) = (A \cdot B) \cdot C$
- c./  $A \cdot (B+C) = A \cdot B + A \cdot C; \quad A+B \cdot C = (A+B) \cdot (A+C)$
- d./  $A+A = A \quad A \cdot A = A$
- e./  $\overline{\overline{A}} = A$
- f./  $A+\emptyset = A; \quad A \cdot \emptyset = \emptyset$
- g./  $A+I = I; \quad A \cdot I = A$
- h./  $A + \bar{A} = I; \quad A \cdot \bar{A} = 0$
- i./  $\overline{A+B} = \bar{A} + \bar{B}, \quad \overline{A \cdot B} = \bar{A} + \bar{B}$

### 1.2.3. A valószínűségekre vonatkozó néhány tulajdonság

- a./  $A$  lehetetlen esemény valószínűsége nulla:  $P(\emptyset) = 0$   
 $A$  biztos esemény valószínűsége egy:  $P(I) = 1$
- b./ Ellentétes események valószínűségeinek összege egy:  $P(A) + P(\bar{A}) = 1$
- c./ Ha  $A$  részhalmaza  $B$ -nek,  $A \subset B$ :  $P(A) \leq P(B)$
- d./ Két tetszőleges esemény összegének valószínűsége:  $P(A+B) = P(A) + P(B) - P(A \cdot B)$   
 Ha  $A$  és  $B$  egymást kizáró események:  $P(A+B) = P(A) + P(B)$ ,  
 mert  $A \cdot B = \emptyset$

#### 1.2.4. Feltételes valószínűség, független események.

Legyenek A és B egy kísérlettel kapcsolatos események. Ha N kísérletet végezve a B esemény n-szer fordul elő, s ezen belül k esetben B-vel együtt A is bekövetkezik, akkor a k/n hányadost az A eseménynek B feltételre vonatkozó feltételes relatív gyakoriságának nevezzük. A feltételes relatív gyakoriság határértéke a feltételes valószínűség.

Jelölje  $P(AB)$  az A és B esemény együttes bekövetkeztének valószínűségét,  $P(B)$  a B esemény bekövetkeztének valószínűségét. Ha  $P(B) > 0$ , akkor az **A esemény B feltétel melletti feltételes valószínűsége**:

$$P(A|B) = \frac{P(AB)}{P(B)} \quad (1.2./1.)$$

Az összefüggés könnyen értelmezhető, ha figyelembe vesszük, hogy a kedvező esetek azok, amikor A és B együtt következett be, aminek valószínűsége  $P(AB)$ . A B-re vonatkozó feltétel teljesülését eredményező összes lehetséges eset valószínűsége  $P(B)$ .

Az A és B események egymástól függetlenek, ha az egyik eseménynek a másikra vonatkozó feltételes valószínűsége megegyezik az esemény feltétel nélküli valószínűségével:

$$P(A|B) = P(A) \quad (1.2./2.)$$

**Független események** együttes bekövetkezésének valószínűsége az 1.2./1. és 1.2./2. összefüggések egybevetése alapján:

$$P(AB) = P(A) \cdot P(B) \quad (1.2./3.)$$

#### 1.2.5. Teljes valószínűség tétele, Bayes tétel.

Ha  $A_1, A_2, \dots, A_n$  teljes eseményrendszer alkotnak és  $P(A_i) > 0$  minden egyes eseményre, akkor egy tetszőleges B esemény valószínűsége a **teljes valószínűség tétele** szerint

$$P(B) = \sum_{j=1}^n P(B|A_j) \cdot P(A_j) \quad (1.2./4.)$$

A **Bayes tétel** az  $A_i$  esemény valószínűségét adja meg a B esemény feltétele mellett.

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{j=1}^n P(B|A_j) \cdot P(A_j)} \quad (1.2./5.)$$

### 1.3. VALÓSZÍNŰSÉGI VÁLTOZÓ

A véletlen tömegjelenségeket alkotó eseményhalmazok elemeihez egy-egy számértéket rendelünk, az így kapott, véletlentől függő változót valószínűségi változónak nevezzük. (Jelölése a továbbiakban:  $\xi$ ). Ha  $\xi$  felvett értékei a számegyenes mentén diszkrét értékek (véges, vagy megszámlálhatóan végtelen halmazt alkotnak), akkor  $\xi$  **diszkrét valószínűségi változó**. Az olyan valószínűségi változót, melynek értékei a számegyenes egy teljes intervallumát (általános esetben a teljes számegyenest) kitöltik, **folytonos valószínűségi változónak** nevezzük.

**Diszkrét valószínűségi változók.** Laboratóriumi gyakorlatban általában ide sorolhatók a számláláson alapuló mérési módszerek (pl. Howard-szám, Petri-csészén megjelenő telepszám, határhígításos módszerrel meghatározott legvalószínűbb élősejtszám, stb), valamint a pontozásos érzékszervi vizsgálatok eredményei.

**Folytonos valószínűségi változók.** Laboratóriumi gyakorlatban általában ide sorolhatók a fizikai és kémiai analitikai eredmények (tömeg, pH, vezetőképesség, koncentráció, stb.)

Szigorú matematikai szempontból tulajdonképpen a fizikai, kémiai analitikai eredmények sem tekinthetők folytonos változónak, hiszen ezen valóban folytonos jellemzők mérésének lehetséges eredményei a műszerek felbontóképessége által meghatározott diszkrét értékekből származnak.

Egy valószínűségi változó folytonosként való kezelhetőségét a gyakorlatban a változó értéktartományának és diszkrét értékközeinek egymáshoz való viszonya dönti el. Ilyen értelemben majdnem mindig folytonos változóként kezelhetők az analitikai eredmények és sok esetben az érzékszervi pontszámok, mikrobaszámok is.

### 1.4. VALÓSZÍNŰSÉGI VÁLTOZÓ ELOSZLÁSFÜGGVÉNYE ÉS SŰRŰSÉGFÜGGVÉNYE

Minden valószínűségi változóra értelmezhető egy  $F(x)$  *eloszlásfüggvény*, amely megadja annak a valószínűségét, hogy a  $\xi$  valószínűségi változó az  $x$  értéknél kisebb.

$$F(x) = P(\xi < x) \quad (1.4/1.)$$

Minél nagyobb  $x$  értéke, annál nagyobb a valószínűsége, hogy  $\xi$  értéke ezt nem éri el. Határesetben:

$$\text{ha } x \rightarrow \infty, \quad \lim F(x) = 1$$

$$\text{ha } x \rightarrow -\infty, \quad \lim F(x) = 0$$

Az  $F(x)$  eloszlásfüggvény monoton nő. Diszkrét valószínűségi változó eloszlásfüggvénye lépcsőzetes alakú.

Az eloszlásfüggvény ismeretében könnyű meghatározni annak a valószínűségét, hogy a  $\xi$  valószínűségi változó az  $a$  és  $b$  értékek közé esik. Feltéve, hogy  $a < b$  :

$$P(a \leq \xi < b) = F(b) - F(a) \quad (1.4/2.)$$

Az  $a$ - $b$  intervallumot tetszőlegesen változtatva, folytonos valószínűségi változóra megadható, hogy  $\xi$  milyen valószínűséggel esik egy adott intervallumba.

Ha az eloszlásfüggvény folytonos és differenciálható, akkor annak deriválásával a **sűrűségfüggvény** meghatározható.

$$\frac{dF}{dx} = f(x) \quad (1.4./3)$$

Az eloszlásfüggvény mindenkori értékét a sűrűségfüggvény integrálja adja meg:

$$F(x) = \int_{-\infty}^x f(x)dx \quad (1.4./4.) \quad \text{illetve:} \quad \int_{-\infty}^{+\infty} f(x)dx = 1 \quad (1.4./5.)$$

Diszkrét valószínűségi változók esetében a sűrűségfüggvény természetesen nem értelmezhető. Ebben az esetben azt a függvényt, amely megadja, hogy a valószínűségi változó milyen ( $p_i$ ) valószínűséggel veszi fel az adott ( $x_i$ ) értéket, **valószínűségi eloszlásnak** nevezzük.

$$p_i = P(\xi = x_i) \quad (1.4./6.)$$

Az eloszlás- és sűrűségfüggvények alakját a konkrét valószínűségi eloszlások tárgyalásánál mutatjuk be.

## 1.5. A VALÓSZÍNŰSÉGI VÁLTOZÓ FŐBB JELLEMZŐI

### 1.5.1. Várható érték (M)

*Diszkrét eloszlású* valószínűségi változó várható értéke:

$$M(\xi) = \sum p_i \cdot x_i \quad (1.5./1.)$$

ahol  $x_i$  a  $\xi$  valószínűségi változó lehetséges értékeit jelenti,  $p_i$  pedig a hozzájuk tartozó  $p_i = P(\xi = x_i)$  valószínűségeket.

*Folytonos eloszlású*, sűrűségfüggvénnyel rendelkező valószínűségi változó várható értéke:

$$M(\xi) = \int_{-\infty}^{+\infty} x \cdot f(x) dx \quad (1.5./2.)$$

A valószínűségi változó várható értékének becslésére szolgál a megfigyelt értékek számtani átlaga.

### A várható értékre vonatkozó néhány tétel

$$a./ \quad \text{Ha } c = \text{konstans,} \quad M(c) = c \quad M(c \cdot \xi) = c \cdot M(\xi)$$

b./	Ha a és b = konstans	$M(a \cdot \xi + b) = a \cdot M(\xi) + b$
c./	Tetszőleges x és y valószínűségi változókra:	$M(x \pm y) = M(x) \pm M(y)$
d./	$M[M(\xi)] = M(\xi)$	
e./	Ha x és y független valószínűségi változók:	$M(xy) = M(x) \cdot M(y)$

### 1.5.2. Szórás (D)

A szórás (D) a valószínűségi változó várható érték körüli ingadozásának a mértékszámát. Meghatározása a szórásnégyzetből ( $D^2$ ) történik.

A *szórásnégyzet* a valószínűségi változó várható értéktől való eltérése négyzetének várható értéke. Matematikai megfogalmazásban sokkal egyszerűbb:

- *Diszkrét eloszlású* valószínűségi változó szórásnégyzete:

$$D^2(\xi) = M [(\xi - M(\xi))^2] = \sum (x_i - M(\xi))^2 p_i \quad (1.5/3.)$$

- *Folytonos eloszlású* valószínűségi változó szórásnégyzete:

$$D^2(\xi) = M [(\xi - M(\xi))^2] = \int_{-\infty}^{+\infty} (x - M(\xi))^2 f(x) dx \quad (1.5/4.)$$

A szórásnégyzetre levezethető tétel, amely számítástechnikailag sokkal könnyebben alkalmazható összefüggéshez vezet:

$$D^2(\xi) = M(\xi^2) - [M(\xi)]^2 \quad (1.5/5.)$$

vagyis a *szórásnégyzet* a valószínűségi változó négyzetének várható értéke és a várható érték négyzetének a különbsége.

A szórás a szórásnégyzet pozitív négyzetgyökeként számítható:

$$D = \sqrt{D^2} \quad (1.5/6)$$

$$D(\xi) = \sqrt{M[\xi - M(\xi)]^2} \quad (1.5/7.)$$

## A szórásnégyzetre vonatkozó néhány tétel

- a./  $D^2(\xi) = M(\xi^2) - M^2(\xi)$
- b./ Ha  $c$ =konstans,  $D^2(c\xi) = c^2 \cdot D^2(\xi)$
- c./ Ha  $x$  és  $y$  független valószínűségi változók:  $D^2(x \pm y) = D^2(x) + D^2(y)$
- d./ Ha  $\bar{x}$   $n$  db. valószínűségi változó számtani közepe:  
$$D^2(\bar{x}) = \frac{D^2(x)}{n}$$
- e./ Ha  $a$  és  $b$ = konstans  $D^2(a\xi + b) = a^2 \cdot D^2(\xi)$

## 1.6. STANDARDIZÁLT VALÓSZÍNŰSÉGI VÁLTOZÓ

Az olyan valószínűségi változót, amelynek várható értéke 0 és szórása 1, standard (standardizált) valószínűségi változónak nevezzük.

Bármely  $x$  valószínűségi változóból standard valószínűségi változó ( $u$ ) hozható létre az alábbi transzformációval:

$$u = \frac{x - M(x)}{D(x)} \quad (1.6./1.)$$

Az összefüggésből a várható értékre és a szórásnégyzetekre vonatkozó tételekkel levezethető:

$$M(u) = 0$$

$$D^2(u) = 1$$

## 1.7. CSEBISEV EGYENLŐTLENSÉG

A Csebisev egyenlőtlenség egy valószínűségi változó várható értéktől való eltérésének valószínűségére ad becslést.

Annak a valószínűsége, hogy a  $\xi$  valószínűségi változó értéke egy adott  $\alpha$  számnál jobban térjen el a várható értékétől, legfeljebb  $D^2/\alpha^2$ .

$$P(|\xi - M(\xi)| > \alpha) \leq D^2/\alpha^2 \quad (1.7./1)$$

ahol  $\alpha = k \cdot D$  és  $k > 1$ .

Az összefüggésből kiolvasható, hogy a várható értéktől való eltérés növeléséhez csökkenő valószínűségek tartoznak.

## 1.8. A VALÓSZÍNŰSÉGI ELOSZLÁS ÁLTALÁNOS JELLEMZŐI

### Momentumok

- Az  $r$ -ed rendű momentum:  $M_r(\xi) = \int_{-\infty}^{+\infty} x^r f(x) dx, \quad r = 1, 2, \dots \quad (1.8./1)$
- Az  $r$ -ed rendű abszolút momentum:  $M_r(|\xi|) = \int_{-\infty}^{+\infty} |x|^r f(x) dx, \quad r = 1, 2, \dots \quad (1.8./2)$
- Az  $r$ -ed rendű centrális momentum:  $\mu_r(\xi) = \int_{-\infty}^{+\infty} [x - M(\xi)]^r f(x) dx, \quad r = 1, 2, \dots \quad (1.8./3)$

Az elsőrendű momentum a valószínűségi változó várható értéke.

A másodrendű centrális momentum a valószínűségi változó szórásnégyzete. Ennek pozitív négyzetgyöke a szórás ( $D$ )

### Ferdeség

Nem szimmetrikus folytonos eloszlás esetében a ferdeségi együttható:  $\gamma_1 = \frac{\mu_3}{D^3}$

### Lapultság

Az eloszlás- ill. a sűrűségfüggvény lapultságát jellemző együttható:

$$\gamma_2 = \frac{\mu_4}{D^4} - 3$$

### Medián

A  $\xi$  valószínűségi változó mediánja ( $m_e$ ) az a szám, amely alatt és felett egyforma (50%) valószínűséggel található értékek, azaz

$$F(m_e) = \int_{-\infty}^{m_e} f(x) dx = \frac{1}{2}$$

### Kvantilis:

A  $\xi$  valószínűségi változó  $q$  kvantilisén azt az  $x_q$  számot értjük, amely alá  $q$ , fölé pedig  $1-q$  valószínűséggel esik változó:

$$F(x_q) = q$$

### Módusz:

*Folytonos eloszlásnál* a valószínűségi változó minden olyan értéke, ahol a sűrűségfüggvénynek helyi maximuma van.

*Diszkrét eloszlásnál* a valószínűségi változó azon értéke, ahol a  $p_i$  valószínűségnek a környezetében lévő valószínűségekhez képest maximuma van.

## 1.9. A VALÓSZÍNŰSÉGI ELOSZLÁS PARAMÉTEREINEK BECSLÉSE

A valószínűségi változó ismeretlen paramétereit a mintából számított értékekkel *becsüljük*. Ilyen értelemben mintának számítanak a kísérleti eredmények is. A becslés (a becslési eljárás eredménye) és a paraméter valódi értéke között általában eltérés van.

A valódi érték és a becslés közötti eltérés egyik része abból ered, hogy a becslés maga is valószínűségi változó, ezért értékét véletlen ingadozások befolyásolják. Ez a fajta *véletlen eltérés* a *konfidencia számítások* segítségével behatárolható és a számításhoz felhasznált adatok számának növelésével csökkenthető.

Az eltérés nem véletlenszerű része a **torzítás**. A torzítás forrásai az alábbiak lehetnek:

- Nem megfelelő a mintavétel, (kísérleti elrendezés), a minták, (kísérleti eredmények) nem reprezentálják a vizsgált sokaságot (a mintavétel torzított).
- Nem megfelelő a vizsgálati eljárás (a vizsgálati módszer torzított).
- Nem megfelelő számítási eljárást használunk a paraméter becslésére (a becslés matematikai eljárása torzított).

A torzítás matematikailag közvetlenül nem ismerhető fel az eredményekből, csupán szakmai megfontolások alapján gyanítható. *A torzítás a vizsgálati adatok számának növelésével nem csökkenthető.*

- A mintavételi torzítás a véletlen mintavételre vonatkozó szabályok betartásával, illetve megfelelő kísérlettervezéssel kerülhető el.
- A vizsgálati módszerek torzítását más módszerekkel való összehasonlítással, valamint azonos mintákkal végzett különböző laboratóriumok közötti összehasonlító vizsgálatokkal lehet megállapítani. Ennek eredményeként a torzított eredményt adó módszerek kiszűrhetők, vagy a torzítás forrása feltárható.
- A nem megfelelő matematikai módszer alkalmazásából eredő torzítás a megfelelő becslési eljárás kiválasztásával kerülhető el. Ez sokszor komoly matematikai megfontolásokat igényel.

Minden becslés valószínűségi változó, s ennek megfelelően eloszlása és eloszlásfüggvénye van. A matematikai-statisztikában alkalmazott legfontosabb valószínűségi eloszlásokat a következő, 2. fejezet tárgyalja.



## 2. VALÓSZÍNŰSÉGI ELOSZLÁSOK

A valószínűségi eloszlások tárgyalásánál a kombinatorikai levezetéseket nem ismertetjük, azok a megadott szakirodalmi forrásokban részletesen megtalálhatók.

A valószínűségi eloszlások közül azokkal foglalkozunk, amelyeknek a matematikai-statisztikai számítások, illetve a minőségellenőrzéssel kapcsolatosan jelentőségük van, de emellett tárgykörünkhöz kapcsolódó néhány speciális eloszlásra is utalunk.

A valószínűségi eloszlások várható értékének és szórásának az 1.5. fejezetben ismertetett összefüggések alapján történő levezetéseit könyvünk nem tárgyalja, azok az irodalomjegyzékben szereplő tankönyvekben megtalálhatók.

### 2.1. DISZKRÉT ELOSZLÁSOK

#### 2.1.1. Egyenletes eloszlás

Ha a  $\xi$  valószínűségi változó  $n$  különböző értékeket egyforma valószínűséggel vehet fel, akkor a  $\xi$  egyenletes eloszlású diszkrét valószínűségi változó.

Valószínűségi függvény:  $P(\xi=x_i) = p_i = 1/n$

Várható érték:  $M(\xi) = \frac{1}{n} \sum_{i=1}^n x_i$  (2.1./1.)

Szórásnégyzet:  $D^2(\xi) = \frac{1}{n} \cdot \sum_{i=1}^n x_i^2 - \left[ \frac{1}{n} \cdot \sum_{i=1}^n x_i \right]^2$  (2.1./2)

#### *Számítási példa egyenletes eloszlásra*

Mi a valószínűsége annak, hogy egy 25 elemű sorszámozott tételből egy bizonyos sorszámú elemet veszünk ki. Feltételezzük, hogy a mintavétel véletlenszerű.

*Megoldás:* Bármely elem kivételének azonos a valószínűsége:

$$p_i = 1/n = 1/25 = \mathbf{0.04}$$

### 2.1.2. Binomiális (Bernoulli) eloszlás

Ha egy kísérletnek két lehetséges kimenetele van, és ezek kölcsönösen kizárják egymást (pl. megfelel, nem felel meg), akkor ezek teljes eseményrendszert alkotnak. Jelöljük A-val és  $\bar{A}$ -vel az alternatív eseményeket, ekkor

$$A + \bar{A} = I$$

Ha a  $P(A) = p$ , és  $P(\bar{A}) = q$  jelölést használjuk, akkor

$$p + q = 1 \quad (2.1./3.)$$

Ismételjük meg n-szer a kísérletet és számítsuk ki annak a valószínűségét, hogy az A esemény k-szor következik be, míg  $\bar{A}$  (n-k)-szor.

Annak a valószínűsége, hogy az  $i_1$  és  $i_2, \dots$  és  $i_k$ -adik kísérletnél A bekövetkezik, és a többinél A nem következik be, tekintettel arra, hogy ezek az események függetlenek egymástól:  $p^k \cdot q^{n-k}$ .

n kísérletből k eset kiválasztásának lehetséges száma (1.1./5. összefüggés szerint)  $\binom{n}{k}$ .

Annak a valószínűségét, hogy n kísérletből az A esemény k-szor következik be, az események lehetséges számának és a bekövetkezés valószínűségének a szorzata adja meg:

$$P_n(k) = \binom{n}{k} \cdot p^k \cdot q^{n-k} \quad (2.1./4.)$$

Ha  $\xi$  valószínűségi változó azon kísérletek száma, amelyeknél az A esemény bekövetkezett, akkor a 2.1./4. összefüggéssel leírt  $P_n(k) = P_n(\xi=k)$  valószínűségek diszkrét eloszlást adnak, amelyet *Bernoulli*, vagy *binomiális* eloszlásnak nevezünk. A binomiális elnevezést az eloszlás az alábbi összefüggés miatt kapta:

$$\sum_{k=0}^n P_n(\xi = k) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1 \quad (2.1./5.)$$

A  $P_n(\xi=k)$  valószínűségek a  $p+q$  binom n-edik hatványának a tagjai.

**Valószínűségi függvény:**

$$p_k = P(\xi = k) = \binom{n}{k} p^k q^{n-k} \quad (k=0, 1, \dots, n) \quad (2.1./6.)$$

**Eloszlásfüggvény:**

$$F(\xi) = \sum_{k < x} p_k = \begin{cases} 0 & \text{ha } x \leq 0 \\ \sum_{k < x} \binom{n}{k} p^k q^{n-k} & \text{ha } 0 < x \leq n \\ 1 & \text{ha } x > n \end{cases} \quad (2.1./7.)$$

A binomiális eloszlás két paramétere:  $p$  és  $n$ , amelyek az eloszlást egyértelműen meghatározzák.

A binomiális eloszlás várható értéke és szórása:

$$\mu = n \cdot p \quad (2.1./8.)$$

$$D = \sqrt{n \cdot p \cdot (1 - p)} \quad (2.1./9.)$$

A binomiális eloszlás fenti levezetése szigorúan véve csak akkor teljesül, ha az egyes mintavételek során a kivett mintaelemek száma (k) nem befolyásolja az A esemény bekövetkeztének valószínűségét (p). Ez általában teljesül, ha a mintaelemek száma (n) elhanyagolható az alapsokaság elemeinek számához viszonyítva, vagy ha a kivett mintaelemet mindig visszatesszük a következő elem kivétele előtt.

Minden olyan esetben, amikor egy esemény bekövetkezését vizsgáljuk, az esemény bekövetkezésének száma, mint valószínűségi változó, binomiális eloszlást követ. A binomiális eloszlásnak rendkívül fontos szerepe van a minőségellenőrzési mintavételi tervek kialakításában.

Kellően nagy mintaszám esetén a binomiális eloszlás jól közelíthető a  $\mu = np$  várható értékű és  $D = \sqrt{n \cdot p \cdot (1 - p)}$  szórású normális eloszlással.

### ***Számítási példák binomiális eloszlásra***

**1. Példa** Mi a valószínűsége annak, hogy egy 10% kifogásolt minőségű terméket tartalmazó tételből 5 elemű mintát kivéve, abban

- a./ nem találunk hibás elemet
- b./ 2 db. hibás elemet találunk
- c./ a hibás elemek száma nem haladja meg a 2-t.

Számítsuk ki az eloszlás várható értékét és szórását.

Megoldás

A hibás elem előfordulási valószínűsége:  $p = 0,1$        $(1-p) = 0,9$

A mintaelemek száma:  $n = 5$

A hibás elemek megengedett száma:  $k$

Alkalmazva a binomiális eloszlásra vonatkozó 2.1./4. összefüggést.

$$P_n(k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

a./  $k = 0$        $P_5(\xi=0) = \binom{5}{0} \cdot 0,1^0 \cdot 0,9^5 = 1 \cdot 1 \cdot 0,9^5 = \mathbf{0,5905}$

b./  $k = 2$        $P_5(\xi=2) = \binom{5}{2} \cdot 0,1^2 \cdot 0,9^3 = \frac{5 \cdot 4}{1 \cdot 2} \cdot 0,01 \cdot 0,7290 = \mathbf{0,0729}$

c./  $k \leq 2$        $\sum_{k=0}^2 P_n(\xi = k) = \sum_{k=0}^2 \binom{5}{k} p^k q^{n-k}$

$$P_5(\xi=1) = \binom{5}{1} \cdot 0,1^1 \cdot 0,9^4 = 5 \cdot 0,1 \cdot 0,9^4 = \mathbf{0,3281}$$

$$P_5(\xi \leq 2) = P_5(\xi=0) + P_5(\xi=1) + P_5(\xi=2) = 0,5905 + 0,3281 + 0,0729 = \mathbf{0,9915}$$

**Várható érték:**  $\mu = n \cdot p = 5 \cdot 0,1 = \mathbf{0,5}$

**Szórás:**  $D = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{5 \cdot 0,1 \cdot (1-0,1)} = \sqrt{0,45} = \mathbf{0,6708}$

**2. Példa.** Az élelmiszerek mikrobiológiai minősítése során patogén mikroorganizmusok jelenléte nem megengedett. Tételminősítés során 10 elemű mintát vesznek és minden egyes elemet megvizsgálják. Megfelelőnek minősítik a tételt, ha patogén mikroba jelenléte egy mintaelemből sem mutatható ki. Mi a valószínűsége annak, hogy egy 20%-ban fertőzött tételt a fenti eljárás elfogadhatónak minősít.

Megoldás

A hibás elem előfordulási valószínűsége:  $p = 0,2$        $(1-p) = 0,8$

A mintaelemek száma:  $n = 10$

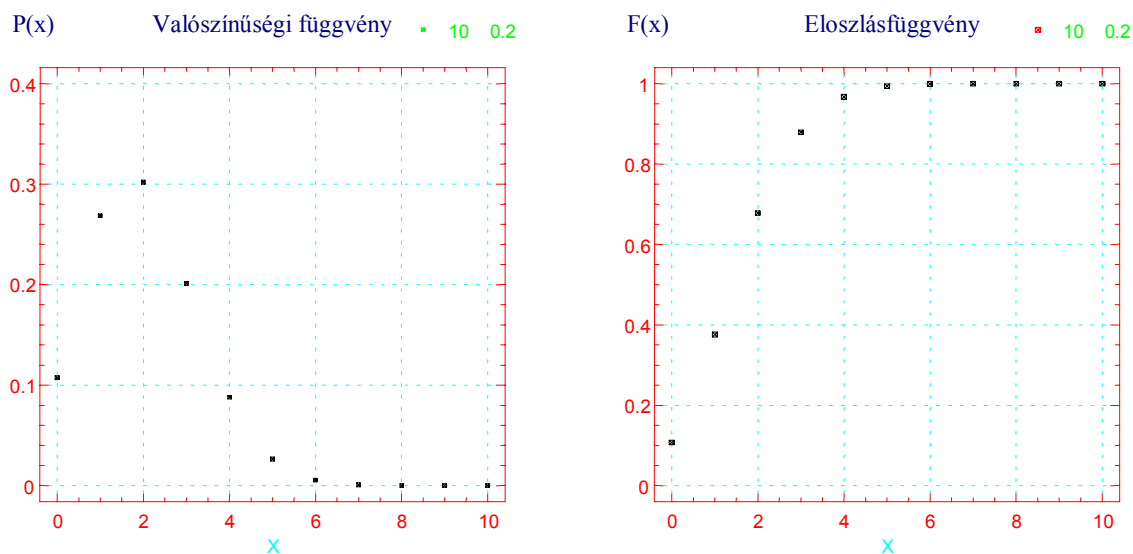
A hibás elemek megengedett száma:  $k = 0$

Alkalmazva a binomiális eloszlásra vonatkozó 2.1./4. összefüggést.

$$P_{10}(\xi=0) = \binom{10}{0} \cdot 0,2^0 \cdot 0,8^{10} = 1 \cdot 1 \cdot 0,8^{10} = \mathbf{0,1074}$$

A hibás minősítés valószínűsége meghaladja a 10%-ot !!

A fenti példában szereplő binomiális eloszlás valószínűségi függvényét és eloszlásfüggvényét a 2.1./1 ábra szemlélteti.



**2.1./1. ábra.** Binomiális eloszlás (n=10, p=0,20) valószínűségi- és eloszlásfüggvénye.

### 2.1.3. Hipergeometrikus eloszlás

Azokban az esetekben, amikor az  $N$  elemű alapsokaságból visszatevés nélkül végezzük a mintavételt, annak a valószínűsége, hogy az  $n$  elemű mintába  $k$  db. nem megfelelő kerül *hipergeometrikus eloszlást* követ.

Ha az alapsokaság száma  $N$ , amelyben a selejtarány  $p$ , akkor az egész sokaságban  $N \cdot p = M$  db. nem megfelelő egyed van.

$$P(\xi = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \quad 2.1./10.$$

A hipergeometrikus eloszlás határértéke, ha  $N \rightarrow \infty$  és  $p = M/N$  a binomiális eloszlás.  $N \geq 50 \cdot n$  viszony teljesülése esetén a hipergeometrikus eloszlás már binomiális eloszlással helyettesíthető a számításokban.

A hipergeometrikus eloszlás várható értéke és szórásnégyzete:

$$\mu = \frac{nM}{N} = n \cdot p \quad 2.1./11.$$

$$D^2 = n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1} \quad 2.1./12.$$

A hipergeometrikus eloszlást egyértelműen meghatározó három paraméter:  $N$ ,  $n$ ,  $p$

A hipergeometrikus eloszlás várható értéke teljesen megegyezik a binomiális eloszlás várható értékével. A szórásnégyzetre vonatkozó 2.1./12. kifejezés pedig  $N$  növelésével alulról közelíti a binomiális eloszlás szórásnégyzetét. Egyetlen mintaelem ( $n=1$ ) kivétele esetén a két eloszlás szórása megegyezik. (Egyetlen mintaelem vizsgálatakor nincs jelentősége a tétel nagyságának, illetve a visszatevésnek, ilyenkor a binomiális eloszlás minden esetben alkalmazható.)

A hipergeometrikus eloszlásnak a minőségellenőrzésben a kis tételek ( $N < 100$ ) mintavételi, illetve minősítési terveinek kialakításában van szerepe.

### 2.1.4. Poisson eloszlás

A laboratóriumi gyakorlatban nagyon sokszor előfordul, hogy egyes eseményeknek egy időtartamon, térfogaton, vagy felületen belüli bekövetkezési gyakoriságát vizsgáljuk. Példa erre a radioaktív bomlások időegység alatti száma, a Bürker kamrás mikroszkópos sejtszámlálás, a Petri csészéken megjelenő telepek száma, stb. Ezekben az esetekben a vizsgált valószínűségi változó (időegység alatti beütésszám, látómezőnkénti sejtszám, Petri csészénkénti telepszám, stb.) *Poisson eloszlást* követ. Az eloszlást a  $k$  esemény  $n$  megfigyelés alatti átlagos értéke ( $\lambda$ ) jellemzi.

A Poisson eloszlás a binomiális eloszlás határeset, ha  $n \rightarrow \infty$ , miközben  $n \cdot p = \lambda$  konstans marad. A gyakorlatban, ha  $p < 0,1$  és  $n > 20$ , a binomiális eloszlás helyettesíthető a Poisson eloszlással.

Annak a valószínűsége, hogy a  $\xi$  változó a  $k$  értéket veszi fel:

Valószínűségi függvény: 
$$P(\xi = k) = p_k = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \quad (2.1./13.)$$

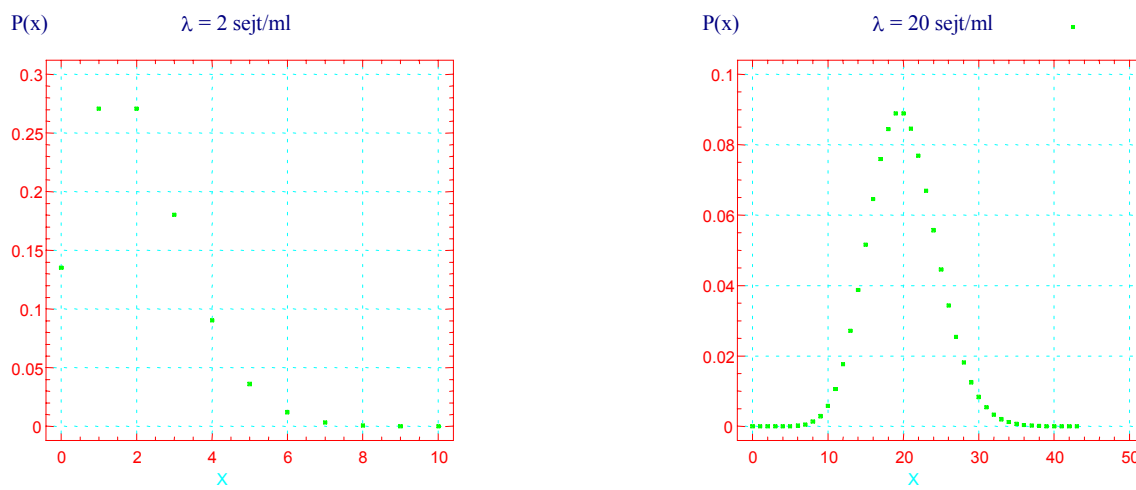
Várható érték: 
$$\mu = \lambda \quad (2.1./14.)$$

Szórásnégyzet: 
$$D^2 = \lambda \quad (2.1./15.)$$

A Poisson eloszlás egyetlen paramétere:  $\lambda$

### Számítási példa Poisson eloszlásra

Lemezöntés elősejtszám meghatározásánál a törzsszuszpenzió elősejtszáma 20 sejt/ml. Hasonlítsuk össze a törzsszuszpenzióból és annak tízszeres hígításából végzett 1 ml minta várható elősejtszám eloszlását. (Feltételezzük, hogy a hígítás tökéletes volt.)



**2.1./2. ábra.** Telepképző egységek elméletileg várható Poisson eloszlása

Összehasonlítva a két eloszlást, jól látható, hogy  $\lambda = 2$  várható érték esetén a sejtszám eloszlása erősen aszimmetrikus, míg  $\lambda = 20$  várható értéknél már igen jó közelítéssel szimmetrikus.

Általában  $\lambda = k/n > 15$  értékeknél a Poisson eloszlás már szimmetrikussá válik és helyettesíthető egy  $\mu = \lambda$  várható értékű és  $D = \sqrt{\lambda}$  szórású normális eloszlással.

## 2.2. FOLYTONOS ELOSZLÁSOK

### 2.2.1. Normális eloszlás

A matematikai-statisztikában elméletileg és gyakorlatilag is egyaránt legfontosabb eloszlás a *normális*-, vagy *Gauss-eloszlás*. A haranggörbe alakú eloszlás jelentőségét az alábbi törvényszerűségeknek köszönheti:

- A véletlen hibák a legtöbb esetben normális eloszlást követnek (innen ered az eloszlás esetenkénti hibatörvény, illetve hibaeloszlás elnevezése).
- A normális eloszlású sokaságból származó minták eloszlása is normális.
- A **centrális határeloszlás tétele** szerint *nagy számú független valószínűségi változó összege közelítőleg normális eloszlású, feltéve hogy az összeg minden egyes tagjának ingadozása kicsi az egész összeg ingadozásához képest*. Ez a gyakorlatban azt jelenti, hogy már három minta átlaga is jó közelítéssel normális eloszlást eredményez, még akkor is, ha az eredeti eloszlás jelentősen eltér a normálistól (pl. kétpupú eloszlás). A centrális határeloszlás tétele következtében a minta elemszám növelésével az eredetileg binomiális és Poisson eloszlású sokaságokra is alkalmazhatóak a normális eloszlásra kidolgozott statisztikai próbák.

**Sűrűségfüggvény:**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} ; \quad (-\infty < x < +\infty) \quad (2.2./1.)$$

**Eloszlásfüggvény:**

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (2.2./2.)$$

**Várható érték:**  $M(X) = \mu$

**Szórásnégyzet:**  $D^2(x) = \sigma^2$

A normális eloszlás két paramétere:  $\mu$  és  $\sigma$ , amelyek az eloszlást egyértelműen meghatározzák.

A normális eloszlás központi jelentősége miatt a statisztikai próbák egységes alkalmazhatósága érdekében bevezették a **standardizált normális eloszlást**, amelynek változója:

$$u = \frac{x - \mu}{\sigma} \quad (2.2./3.)$$

A (2.2./3.) összefüggés alapján bármely  $\mu$  és  $\sigma$  paraméterű normális eloszlás standardizálható, így a matematikai-statisztikai számításokhoz elegendő a standard normális eloszlás táblázatait használni.

**A standardizált normális eloszlás sűrűségfüggvénye:**

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} ; \quad \varphi(-u) = \varphi(u) \quad (2.2./4.)$$

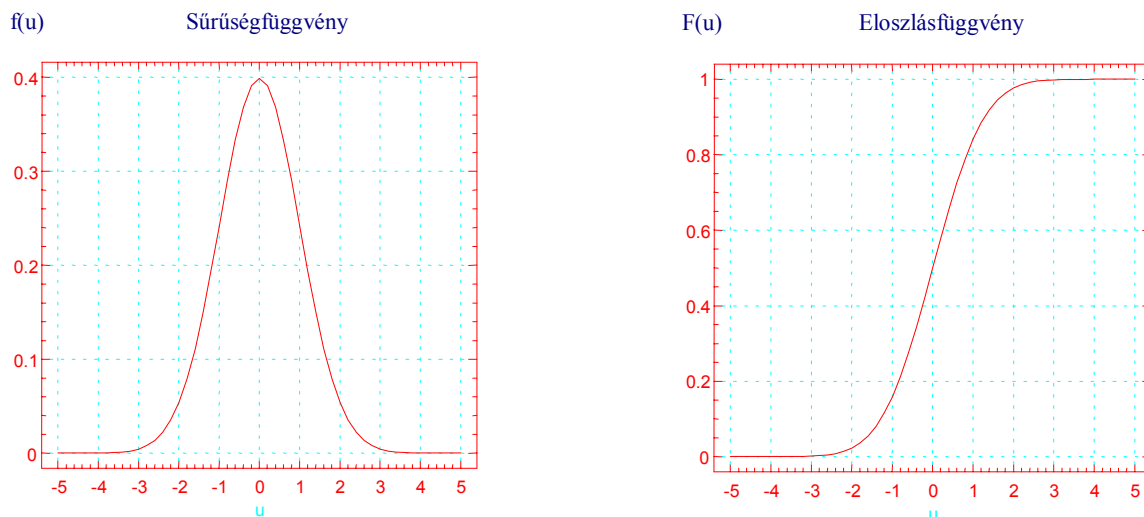
**A standardizált normális eloszlás eloszlásfüggvénye:**

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^u e^{-\frac{t^2}{2}} dt ; \quad \Phi(-u) = 1 - \Phi(u) \quad (2.2./5.)$$

**Várható értéke:**  $\mu = 0$

**Szórása:**  $\sigma = 1$

A standardizált normális eloszlás sűrűség- és eloszlásfüggvényét a 2.2./1. ábra szemlélteti.



**2.2./1. ábra.** Standardizált normális eloszlás sűrűség- és eloszlás függvénye



### 2.2.2. Student eloszlás

A Student eloszlás az átlagérték- és konfidencia-intervallumok becslésekor játszik igen fontos szerepet azokban az esetekben, amikor a vizsgált sokaság szórását és várható értékét is a mintából becsüljük.

Az  $n$  számú ( $m = n-1$  szabadsági fokú  $x_0, x_1, \dots, x_m$ ) független,  $0$  várható értékű,  $\sigma$  szórású normális eloszlású valószínűségi változóból képzett

$$t = \frac{x_0}{\sqrt{\frac{1}{m}(x_1^2 + x_2^2 + \dots + x_m^2)}} \quad (2.2./6.)$$

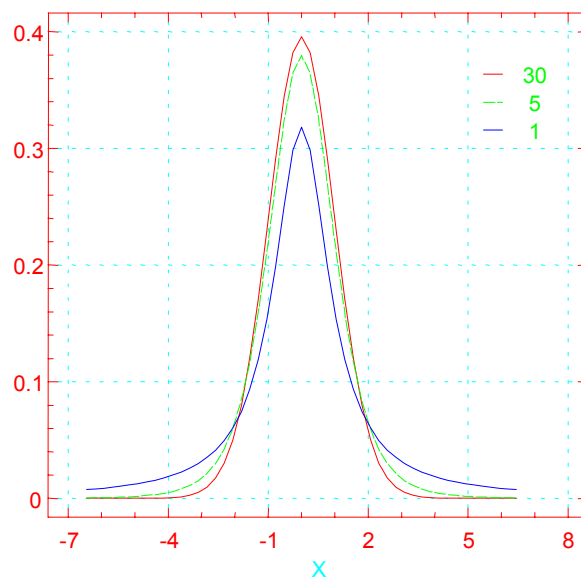
valószínűségi változók  $m$  szabadsági fokú Student ( $t$ ) eloszlást követnek.

**Az eloszlás sűrűségfüggvénye:**

$$f(t) = \frac{1}{\sqrt{m\pi}} \frac{\left(\frac{m-1}{2}\right)!}{\left(\frac{m-2}{2}\right)! \left(1 + \frac{t^2}{m}\right)^{\frac{m+1}{2}}} \quad (2.2./7.)$$

**Várható értéke:**  $\mu = 0$

**Szórásnégyzete:**  $D^2 = m/(m-2)$



**2.2./2. ábra.** Student féle  $t$  eloszlás 1, 5 és 30-as szabadsági fokhoz tartozó sűrűségfüggvényei.

A sűrűségfüggvény  $t=0$ -ra nézve szimmetrikus, alakja hasonló a standardizált normális eloszláshoz. Várható értéke 0, szórásnégyzete a szabadsági fok ( $m$ ) növelésével felülről közelít 1-hez.

Ha  $m > 30$ , akkor a  $t$ -eloszlás igen jól közelíthető a standardizált normális eloszlással:  $t_p \approx u_p$ .

A Student féle  $t$  eloszlást szemléltető 2.2./2. ábrán jól látható, hogy az 5-ös szabadsági fokhoz tartozó sűrűség-függvény már viszonylag jól közelíti a 30-as szabadsági fokhoz tartozó függvényt.

### 2.2.3. $\chi^2$ eloszlás

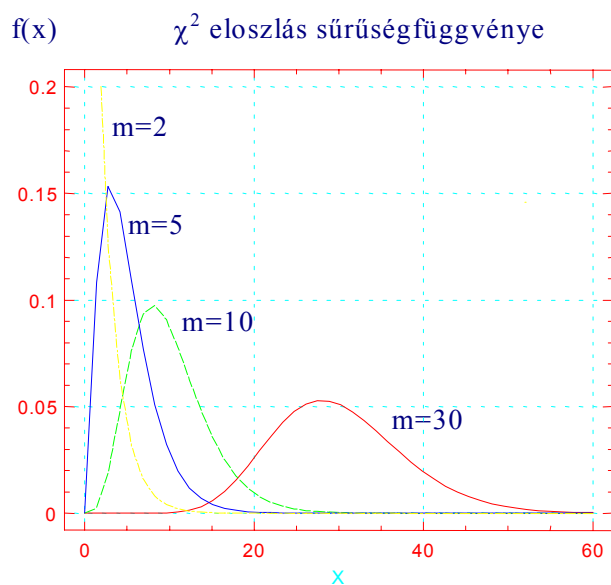
A  $\chi^2$  eloszlás a szórásnégyzetekkel kapcsolatos statisztikai próbáknál, valamint illeszkedés-vizsgálatoknál játszik szerepet

Ha  $x_1, x_2, \dots, x_m$  független, normális eloszlású valószínűségi változók, és  $u_k = (x_k - \zeta_k)/\sigma_k$  standardizáltjaik, akkor a  $\chi^2 = \sum_{k=1}^m u_k^2$  valószínűségi változó  $m$  szabadsági fokú  $\chi^2$  eloszlást követ, melynek sűrűségfüggvénye:

$$f(\chi^2) = \frac{1}{\left(\frac{m-2}{2}\right)! 2^{\frac{m}{2}}} (\chi^2)^{\frac{m-2}{2}} e^{-\frac{\chi^2}{2}} \quad (2.2./8.)$$

Várható értéke:  $\mu = m$

Szórásnégyzete:  $D^2 = 2m$



**2.2./3. ábra.**  $\chi^2$  eloszlás sűrűségfüggvénye ( $m = 2, 5, 10, 30$ )

Ha a  $\chi^2$  eloszlás szabadsági foka nő, az eloszlás  $m$  várható értékű és  $2m$  szórásnégyzetű normális eloszláshoz közelít, amit a 2.2./3. ábra szemléltet.

## 2.2.4. F eloszlás

Az F eloszlás szórásnégyzetek összehasonlításában, illetve az erre visszavezethető statisztikai próbákban (pl. varianciaanalízis) játszik fontos szerepet.

Két  $\chi^2$  eloszlású  $m$ , ill.  $n$  szabadsági fokú független valószínűségi változóból képzett

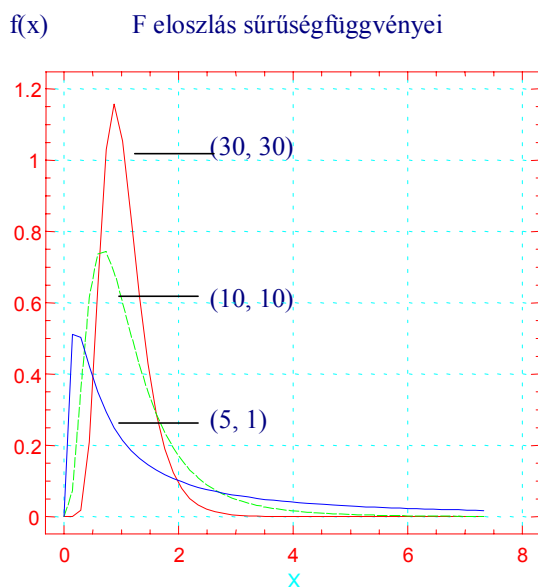
$$F = \frac{\frac{1}{m}(x_1^2 + x_2^2 + \dots + x_m^2)}{\frac{1}{n}(y_1^2 + y_2^2 + \dots + y_n^2)} \quad (2.2./9.)$$

hányados, valószínűségi változó  $(m, n)$  szabadsági fokú F eloszlást követ, melynek sűrűségfüggvénye:

$$f(F) = \frac{\left(\frac{m+n-2}{2}\right)! F^{\frac{m-2}{2}} m^{\frac{m}{2}} n^{\frac{n}{2}}}{\left(\frac{m-2}{2}\right)! \left(\frac{n-2}{2}\right)! (n+mF)^{\frac{m+n}{2}}} \quad (2.2./10.)$$

Várható értéke:  $\mu = \frac{n}{n-2} \quad (n \geq 3)$       Szórásnégyzete:  $D^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad (n \geq 5)$

A különböző szabadsági fokokhoz tartozó sűrűségfüggvényeket a 2.2./4. ábra szemlélteti



**2.2./4. ábra.** F eloszlás sűrűségfüggvényei ( $m, n = 30, 30; 10, 10; 5, 1$ )

### 2.2.5. Exponenciális eloszlás

Az exponenciális eloszlás a higitási sorokon alapuló élősejtszám-meghatározási módszerekben, eltarthatósági vizsgálatokban, minőségellenőrzési jelleggörbék kialakításában játszik fontos szerepet.

Az exponenciális eloszlás sűrűség- és eloszlásfüggvénye (csak pozitív  $x$  értékekre értelmezve):

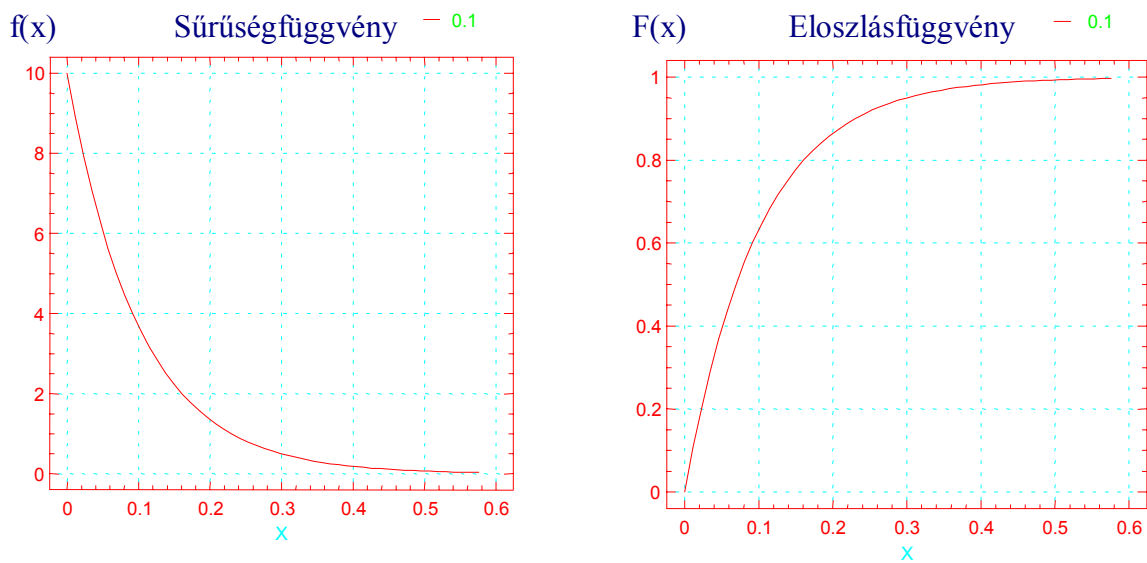
$$\text{Sűrűségfüggvény: } f(x) = \lambda e^{-\lambda x}$$

$$\text{Eloszlásfüggvény: } F(x) = 1 - e^{-\lambda x}$$

Az exponenciális eloszlás várható értéke és szórása megegyezik.

$$\text{Várható érték: } \mu = 1/\lambda$$

$$\text{Szórás } D = 1/\lambda$$



**2.2.5. ábra.** Exponenciális eloszlás sűrűség- és eloszlásfüggvénye ( $\lambda = 0,1$ )

### 3. STATISZTIKAI BECSLÉSEK

A statisztika empirikus adatok összegzése alapján a vizsgált sokaságra vonatkozóan következtetéseket von le és feltevéseket ellenőriz (numerikus becslések, összefüggésekre vonatkozó feltevések, előrejelzések, döntések). Ennek megfelelően a statisztika feladata:

1. A megfigyelt sokaság statisztikai leírása, statisztikai jellemzők kiszámítása.
2. A valószínűségekre alapuló megfelelő matematikai modell (elméleti eloszlás) feltételezése, (vagy felállítása) és ellenőrzése illeszkedésvizsgálattal.
3. A feltételezett valószínűségi modell alapján következtetések levonása, előrejelzések és döntések meghozatala. Hipotézisvizsgálatok.

A megfigyelt sokaság *statisztikai leírása* (1.) a gyakorlatban a mintából meghatározott néhány jellemző (átlagérték, relatív gyakoriság, szórás, terjedelem, stb.) meghatározását jelenti. Ezek a paraméterek a továbbiakban alapul szolgálhatnak a megfigyelt sokaság valószínűségi eloszlásának (*matematikai modell*) feltételezésére, valamint a feltételezés helyességének ellenőrzésére (2.). Amennyiben a mintából meghatározott jellemzők alapján elvégzett *statisztikai próbák* nem mondanak ellent a megmintázott sokaságra vonatkozó hipotéziseinknek, a feltételezett elméleti eloszlásra vonatkozó matematikai apparátus felhasználásával elvégezhetjük a szükséges statisztikai próbákat az alapsokaságra, vagy összehasonlítható alapsokaságokra vonatkozó feltételezéseink ellenőrzésére, döntéseink és előrejelzéseink meghozatalára (3.).

A gyakorlatban nagyon sok esetben előzetes információk alapján, vagy elméleti megfontolásokból kiindulva ismertnek tekintjük a megfigyelt sokaság eloszlását, ezért a statisztikai kiértékelés során a 2. pontban foglalt lépések kihagyásra kerülnek. Az eloszlásra vonatkozó hipotézis helytelen volta azonban nagyon súlyos döntési hibákhoz vezethet.

A megfigyelt sokaság statisztikai jellemzőit mindig a minta alapján becsüljük, s ennek következtében - mint ahogy arra az 1.9. fejezetben utaltunk -, a becslési eljárás eredménye és a paraméter valódi értéke között eltérés lehetséges.

Az alapsokaság (általunk ismeretlen) statisztikai jellemzői adottak. (Egy konzerv készítmény grammonkénti aerob spóraszám, Salmonella jelenlét/hiány egy tejpör-tételben, stb.) Ezeket a valós és konkrét értékű jellemzőket becsüljük a mintából meghatározott értékek alapján. A becslés eredménye egy adott mintából meghatározva szintén konkrét érték, de ez az érték ismételt mintavétel esetén, vagy egyes mintaelemeket kihagyva, illetve további mintaelemeket bevonva az értékelésbe, a vizsgált jellemző valószínűségi változó volta miatt ingadozik.

A matematikai-statisztikai kiértékelés célja soha nem lehet (mert gyakorlatilag nem is lehetséges) a becsült jellemző alapsokaságon belüli értékének teljes pontosságú meghatározása. Becslési eljárásunk eredményeként csupán azt tudjuk megadni, hogy a vizsgált jellemzőnek az alapsokaságon belüli átlagértéke, szórása, gyakorisága, stb. egy adott valószínűséggel milyen határok között van. Ennek az ún. *konfidencia-intervallumnak* a szélessége viszont a kísérlettervezéstől és értékelési módtól függően (általában a mintaszám növelése révén) csökkenthető.

Több sokaság (pl. különböző táptalajokon, vagy eltérő módszerrel meghatározott mikrobaszámok, különböző gyártási tételek mikrobás fertőzöttsége, stb.) összehasonlításakor általában az egyes sokaságokból vett minták statisztikai jellemzőit hasonlítjuk össze. Ezekben az esetekben kérdésselünk nem az, hogy az egyes sokaságok vizsgált jellemzői megegyeznek-e egymással, hanem az, hogy a köztük lévő eltérés meghaladja-e a véletlen ingadozásból eredő mértéket. A véletlen ingadozás mértékét meghaladó különbségeket *szignifikánsnak* nevezzük.

Általánosságban a minta statisztikai jellemzőinek kiértékelésén alapuló döntésünk helyes megfogalmazása szerint az összehasonlított sokaságok statisztikai jellemzői, vagy egyes statisztikai jellemzők becslt és feltételezett értékei között meghatározott (kiszámított) különbség szignifikáns, vagy nem szignifikáns.

A matematikailag szignifikánsnak bizonyuló eltérések azonban nem jelentenek feltétlenül szakmailag is jelentős különbséget. (Pl. lemezöntéses és felületi szélesztéses mikrobaszám-meghatározási módszerek szórása, és néha átlagértéke között is szignifikáns különbség mutatható ki. Ez a különbség szakmailag csupán akkor jelentős, ha a mikrobaszám egy nagyságrenden belüli változását kívánjuk pontosan nyomonkövetni. Több nagyságrendnyi változás mérésére a két módszer egyformán alkalmazható.)

A vizsgált valószínűségi változó valódi értéke és becslése közötti eltérések forrásairól (véletlen eltérés és torzítás) az 1.9. fejezetben már szóltunk. A továbbiakban csupán a statisztikai becsléssel kapcsolatos elvárásokat ismertetjük

Az alapsokaság ismeretlen (de konkrét,  $a$  értékű) paraméterére vonatkozó statisztikai becsléssel ( $\alpha$ ) szembeni elvárásaink a következőkben foglalhatók össze.

A statisztikai becslés legyen:

- **Torzítatlan.** A becslési eljárás várható értéke a keresett paraméter valódi értéke legyen.  $M(\alpha)=a$ .
- **Hatékony.** Az  $\alpha$  statisztika szórásának  $a$  környezetében minimuma legyen.
- **Konzisztens.** A mintaszám növekedésével a paraméter becslt értéke közelítsen a sokaságon belüli valódi értékhez.
- **Elégséges.** Tartalmazzon minden információt, amely a mintából az ismeretlen  $a$  paraméterre nyerhető.

A statisztikai becslések alapján hozott döntések mindig a sokaságra (és nem a mintára) vonatkoznak és kockázatot rejtenek magukban. Ennek a kockázatnak az eredete, a vizsgált jellemző valószínűségi változó volta. Előfordulhat, hogy az alapsokaság olyan részéből veszünk mintát, amely csak igen kis gyakorisággal szerepel (pl. normális eloszlásnak csak a széléből). Ilyen esetekben a mintából számított becslt értékek alapján a sokaságra vonatkozóan téves következtetést vonhatunk le.

Az alapsokaságra vonatkozó feltételezésünk (*nullhipotézis*,  $H_0$ ) a valóságban vagy fennáll, vagy nem. A nullhipotézisre vonatkozóan a minta alapján hozzuk meg döntésünket, ami vagy helyes, vagy téves. A minta alapján hozott döntések lehetséges eseteit a Döntési táblázatban foglaltuk össze.

**Döntési táblázat.** Becslési eljárás alapján hozott döntések lehetséges esetei.

Döntésünk szerint az alapfeltevés	Az alapfeltevés a valóságban	
	Teljesül	Nem teljesül
Teljesül	Helyes döntés	Hibás döntés $\beta$ <b>másodfajú hiba</b>
Nem teljesül	Hibás döntés $\alpha$ <b>elsőfajú hiba</b>	Helyes döntés

A Döntési táblázatból kitűnik, hogy döntéseink során kétféle hibát követhetünk el: első- és másodfajú hibát

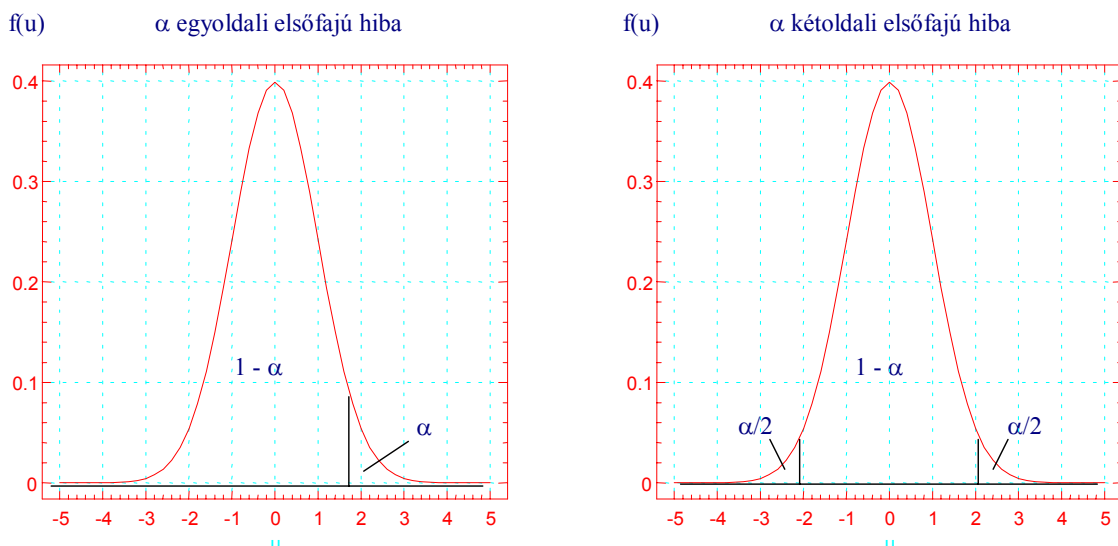
**Elsőfajú hiba  $\alpha$  valószínűséggel** jelenti azt az esetet, amikor döntésünk szerint a nullhipotézis nem teljesül, a valóságban azonban fennáll az alapfeltételezés. (Pl. egy valójában megfelelő tételt hibásnak minősítünk.)

Az elsőfajú hibát a matematikai-statisztikai számítások során előre meg tudjuk határozni, ki tudjuk választani, hogy döntésünket milyen  $P = 1 - \alpha$  biztonsággal kívánjuk meghozni. Ez gyakorlatilag azt jelenti, hogy a vizsgált jellemzőnek csak azt az értéktartományát fogadjuk el, amely  $P = 1 - \alpha$  valószínűséggel az eloszláson belül van. Az eloszláshoz tartozó értékeknek azt az  $\alpha$  töredékét, amely ezen az értékhatáron kívül van, elutasítjuk, nem tekintjük az eloszláshoz tartozónak. Tekintettel arra, hogy a valószínűségi változó sűrűségfüggvénye alatti teljes terület nagysága 1, az elsőfajú hiba mértékét a  $P = 1 - \alpha$  valószínűséget meghaladó terület reprezentálja. Attól függően, hogy az elutasítás az eloszlás mindkét szélére, vagy csak az egyik oldalára vonatkozik, beszélünk kétoldali, vagy egyoldali elsőfajú hibáról.

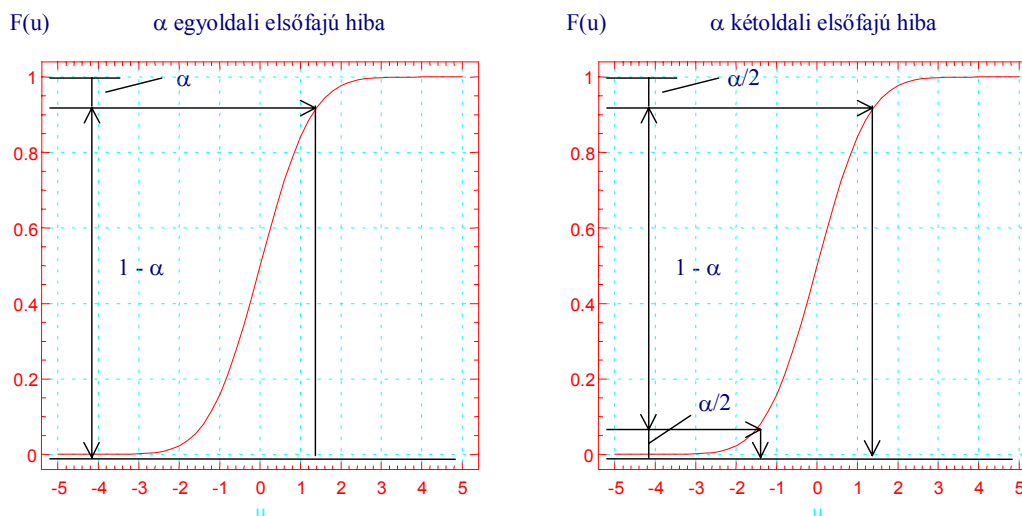
Kétoldali elsőfajú hiba azt jelenti, hogy az eloszlás mindkét széléről elutasítjuk az  $\alpha/2$  területhez tartozó értékeket és elfogadjuk a közéjük eső értéktartományt. Jellemzően ide tartoznak a konfidencia-intervallum számítások, amikor azt állítjuk, hogy a mért jellemző sokaságon belüli értéke  $P = 1 - \alpha$  valószínűséggel a mintában meghatározott érték  $\pm$  (számított) környezetében helyezkedik el. Kétoldali elsőfajú hiba megengedésével az összehasonlítandó sokaságok jellemzői közötti kétirányú eltérések szignifikanciáját vizsgáljuk.

Egyoldali elsőfajú hiba esetén csupán az eloszlás egyik oldalát utasítjuk el. Az elutasított értéktartományhoz tartozik az eloszlási görbe alatti terület  $\alpha$  része. Egyoldali elsőfajú hiba azon döntéseinkre jellemző, melyekben azt vizsgáljuk, hogy egy jellemző becslt értéke a sokaságban meghalad-e (vagy kisebb-e) egy megadott értéknél. Egyoldali elsőfajú hiba esetében csak az egyirányú eltérés szignifikanciáját vizsgáljuk.

Az elsőfajú hibát standardizált normál eloszlás esetén a 3./1. és 3./2. ábrával szemléltetjük.



**3./1. ábra.** Egyoldali és kétoldali elsőfajú hiba sűrűségfüggvényekkel szemléltetve



**3./2. ábra.** Egyoldali és kétoldali elsőfajú hiba eloszlásfüggvényekkel szemlélítve

**Másodfajú hiba  $\beta$  valószínűséggel** jelenti azt az esetet, amikor döntésünk szerint a nullhipotézis teljesül, a valóságban azonban nem áll fenn az alapfeltételezés. (Pl. egy valójában kifogásolandó tételt megfelelőnek minősítünk.) A másodfajú hibát döntéseinkben speciális esetektől eltekintve nem tudjuk előírni és általában ki sem tudjuk számítani.

A kétféle hiba általában együttesen mozog, ha  $\alpha$  nagy lehet, akkor  $\beta$  csökken és fordítva. Együttes csökkentésük csak a mintaelemszám növelésével érhető el.

### 3.1. A MINTA STATISZTIKAI JELLEMZŐI

A matematikai statisztika értelmezése szerint a megfigyelési adatokból számított minden értéket (átlag, szórás, terjedelem, stb.) statisztikának nevezünk. Tekintettel arra, hogy a mintaelemek valószínűségi változók, nyilvánvaló, hogy a belőlük számított statisztikák is valószínűségi változók lesznek, s ennek megfelelően azok eloszlását az alapsokaság eloszlása, az ún. *mintaeloszlás* határozza meg. A mintaeloszlásra vonatkozóan általában rendelkezünk előzetes feltevésekkel (normális-, Poisson-eloszlás, stb.). Amennyiben ilyen ismereteink nincsenek, korrekt kiértékelés csak a tapasztalati mintaeloszlás meghatározása után végezhető el.

#### 3.1.1. Tapasztalati eloszlás meghatározása

Az alapsokaság eloszlását a belőle vett minták tapasztalati eloszlása alapján becsüljük. A tapasztalati eloszlás alakjából következtetünk az alapsokaság-beli valódi eloszlásra, és feltételezésünkre vonatkozóan statisztikai próbákat végzünk.

A tapasztalati eloszlás a meghatározásához felhasznált mintaelemek számának növelésével egyre inkább közelít az alapsokaság valódi eloszlásához. A tapasztalati eloszlás felvételéhez



legalább 100 körüli adat szükséges. A tapasztalati eloszlás meghatározásának a menete a következő.

Véletlenszerű mintavétellel kivesszünk az alapsokaságból lehetőleg minél több (de legalább 100) mintaelemet, amelyből meghatározzuk a vizsgálni kívánt paramétert. (Ilyen értelemben mintavételnek tekinthető egy mérés nagyon sokszori megismétlése is, ahol a mintaelemeket az egyes mérési eredmények jelentik.)

- ◆ Kiválasztjuk a mérési eredmények közül a legnagyobb és legkisebb értéket, ezek különbsége a *terjedelem*.
- ◆ A terjedelmet *értékközökre* osztjuk fel. Az értékközök (intervallumok) számára vonatkozó általános előírás nincs. A gyakorlatban 10-25 azonos szélességű értékközök szokás megadni, a terjedelem és a mérési módszer által szolgáltatott diszkrét értékek egymáshoz való viszonyának figyelembevételével. (Ha egy mérleg csak 0,1 g pontosságú adatokat ad, akkor az 1,5 g szélességű tartományt nincs értelme 15-nél több osztályra felosztani.). Nagy adatszámok esetén az értékközök célszerű száma  $= 2,5 \sqrt[4]{\text{adatszám}}$   
Az értékközök határának kijelölésénél célszerű úgy eljárni, hogy az adatok besorolása egyértelmű legyen. (A határra eső adatokat következetesen vagy a kisebbik, vagy a nagyobbik értékközbe soroljuk.)
- ◆ Meghatározzuk, hogy hány adat tartozik az egyes értékközökbe, (osztályokba), azaz kiszámítjuk az egyes értékközökhöz tartozó gyakoriságot.
- ◆ Megszerkesztjük a gyakoriságokat az értékközök függvényében ábrázoló *gyakorisági hisztogrammot*, vagy a *relatív gyakorisági hisztogrammot*. (A relatív gyakoriság a gyakoriság és az összes adatszám hányadosa.)

Folytonos valószínűségi változó esetében a skála finomításával csökken a gyakoriság. Ennek kiküszöbölése érdekében nem az osztályközönkénti relatív gyakoriságokat, hanem a belőlük származtatott *relatív gyakoriság-sűrűség hisztogrammot* határozzuk meg.

$$\text{Relatív gyakoriság sűrűség} = \text{Relatív gyakoriság} / \text{Osztályszélesség}$$

Könnyen bizonyítható, hogy a relatív gyakoriság-sűrűség hisztogram alatti terület 1.

Ha az értékközök száma túlságosan nagy, a tapasztalati eloszlás véletlen okozta, gyakorlati szempontból jelentéktelen ingadozásai túlzottan érvényesülnek, ha az értékközök száma túl kicsiny, az eloszlás jellege nem mutatkozik kellőképpen.

- ◆ A gyakorisági értékeket összegezve a *kumulatív gyakoriságokat* kapjuk eredményül. Hasonló módon meghatározható a *kumulatív relatív gyakoriság* is. A kumulatív gyakoriságok, vagy relatív gyakoriságokat ábrázolva a hozzájuk tartozó osztályközök függvényében, kapjuk a *tapasztalati eloszlás-függvényt*.

A tapasztalati eloszlás meghatározására ma már általában számítógépes eljárásokat alkalmazunk, amelyek nem csupán a hisztogramok felvételét könnyítik meg, hanem egyúttal az eloszlásra vonatkozó statisztikai próbákat, valamint a teljes matematikai-statisztikai kiértékelést is elvégzik.

A tapasztalati eloszlás vizsgálatát egy mintapéldán keresztül szemléltetjük, amelyre a továbbiakban még többször visszatérünk.

### Számítási példa eloszlásvizsgálatra

Lemezöntéses élősejtszám-meghatározás eredményeinek eloszlását vizsgáltuk pasztörtej mezofil aerob élősejtszámának meghatározására vonatkozóan. Ennek érdekében homogén tejmintából 100 független élősejtszám-meghatározást végeztünk. A meghatározott ml-enkénti élősejtszámokat (N) és azok logaritmusait (lg N) a 3./1. táblázat tartalmazza.

#### **3./1. Táblázat.** Tejminta mezofil aerob élősejtszámai (N) és azok logaritmusai (lg N)

N	lg N	N	lg N	N	lg N	N	lg N
$5.60 \cdot 10^3$	3.75	$1.31 \cdot 10^4$	4.12	$1.78 \cdot 10^4$	4.25	$2.18 \cdot 10^4$	4.34
$6.60 \cdot 10^3$	3.82	$1.45 \cdot 10^4$	4.16	$1.82 \cdot 10^4$	4.26	$2.63 \cdot 10^4$	4.42
$6.90 \cdot 10^3$	3.84	$1.51 \cdot 10^4$	4.18	$1.66 \cdot 10^4$	4.22	$2.70 \cdot 10^4$	4.43
$7.20 \cdot 10^3$	3.86	$1.44 \cdot 10^4$	4.16	$1.67 \cdot 10^4$	4.22	$2.82 \cdot 10^4$	4.45
$8.30 \cdot 10^3$	3.92	$1.41 \cdot 10^4$	4.15	$1.90 \cdot 10^4$	4.28	$2.88 \cdot 10^4$	4.46
$8.70 \cdot 10^3$	3.94	$1.51 \cdot 10^4$	4.18	$1.86 \cdot 10^4$	4.27	$2.76 \cdot 10^4$	4.44
$9.10 \cdot 10^3$	3.96	$1.48 \cdot 10^4$	4.17	$1.91 \cdot 10^4$	4.28	$3.02 \cdot 10^4$	4.48
$8.90 \cdot 10^3$	3.95	$1.45 \cdot 10^4$	4.16	$1.82 \cdot 10^4$	4.26	$2.81 \cdot 10^4$	4.45
$8.70 \cdot 10^3$	3.94	$1.38 \cdot 10^4$	4.14	$1.86 \cdot 10^4$	4.27	$3.00 \cdot 10^4$	4.48
$9.10 \cdot 10^3$	3.96	$1.31 \cdot 10^4$	4.12	$1.73 \cdot 10^4$	4.24	$1.44 \cdot 10^4$	4.16
$9.10 \cdot 10^3$	3.96	$1.45 \cdot 10^4$	4.16	$2.14 \cdot 10^4$	4.33	$2.63 \cdot 10^4$	4.42
$8.70 \cdot 10^3$	3.94	$1.51 \cdot 10^4$	4.18	$2.08 \cdot 10^4$	4.32	$2.69 \cdot 10^4$	4.43
$7.20 \cdot 10^3$	3.86	$1.44 \cdot 10^4$	4.16	$2.24 \cdot 10^4$	4.35	$3.00 \cdot 10^4$	4.48
$1.12 \cdot 10^4$	4.05	$1.31 \cdot 10^4$	4.12	$2.29 \cdot 10^4$	4.36	$3.30 \cdot 10^4$	4.52
$1.10 \cdot 10^4$	4.04	$1.41 \cdot 10^4$	4.15	$2.09 \cdot 10^4$	4.32	$3.63 \cdot 10^4$	4.56
$1.20 \cdot 10^4$	4.08	$1.66 \cdot 10^4$	4.22	$2.19 \cdot 10^4$	4.34	$3.80 \cdot 10^4$	4.58
$1.10 \cdot 10^4$	4.04	$1.82 \cdot 10^4$	4.26	$2.18 \cdot 10^4$	4.34	$3.78 \cdot 10^4$	4.58
$1.20 \cdot 10^4$	4.08	$1.91 \cdot 10^4$	4.28	$2.30 \cdot 10^4$	4.36	$3.38 \cdot 10^4$	4.53
$1.12 \cdot 10^4$	4.05	$1.81 \cdot 10^4$	4.26	$2.40 \cdot 10^4$	4.38	$1.51 \cdot 10^4$	4.18
$1.07 \cdot 10^4$	4.03	$1.70 \cdot 10^4$	4.23	$2.19 \cdot 10^4$	4.34	$3.63 \cdot 10^4$	4.56
$1.15 \cdot 10^4$	4.06	$1.66 \cdot 10^4$	4.22	$2.20 \cdot 10^4$	4.34	$3.63 \cdot 10^4$	4.56
$1.10 \cdot 10^4$	4.04	$1.86 \cdot 10^4$	4.27	$2.14 \cdot 10^4$	4.33	$4.10 \cdot 10^4$	4.62
$1.15 \cdot 10^4$	4.06	$1.77 \cdot 10^4$	4.25	$2.08 \cdot 10^4$	4.32	$4.36 \cdot 10^4$	4.64
$1.05 \cdot 10^4$	4.02	$1.82 \cdot 10^4$	4.26	$2.24 \cdot 10^4$	4.35	$4.78 \cdot 10^4$	4.68
$1.32 \cdot 10^4$	4.12	$1.91 \cdot 10^4$	4.28	$2.30 \cdot 10^4$	4.36	$5.70 \cdot 10^4$	4.76

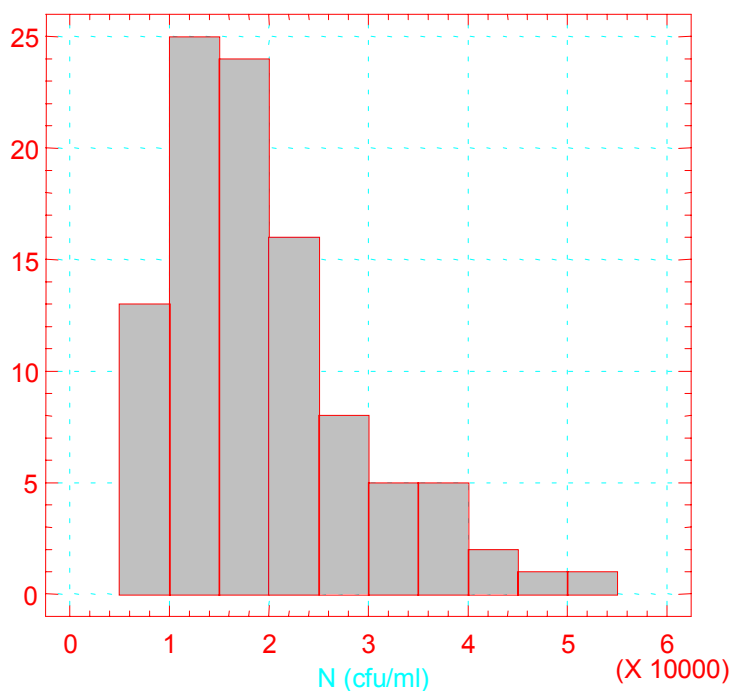
A 3./1. táblázatban összefoglalt adatok osztályba sorolt értékeit a 3./2. és 3./3. táblázatok tartalmazzák. A gyakorisági hisztogramokat a 3./3. és 3./4. ábrán tüntettük fel.

A gyakorisági hisztogramok ábráit összehasonlítva, jól látható, hogy az alapadatok eloszlása erősen aszimmetrikus, a magasabb sejtszámok felé elhúzódozó jelleggel. A logaritmikus transzformáció ezt az eloszlást normális eloszláshoz közelítve, szimmetrikussá teszi. A kétféle tapasztalati eloszlás normálistól való eltérését a továbbiakban  $\chi^2$ -próbával vizsgáljuk.

**3./2. Táblázat.** Mezofil aerob élősejtszám értékeinek osztályba-sorolása

Osztály	Értékköz (x10 <sup>4</sup> )	Érték-közép (x10 <sup>4</sup> )	Gyakoriság	Relatív gyakoriság	Kumulatív gyakoriság	Kumulatív rel. gyak.
1	0.5-1.00	0.75	13	0.13	13	0.13
2	1.01-1.50	1.25	25	0.25	38	0.38
3	1.51-2.00	1.75	24	0.24	62	0.62
4	2.01-2.50	2.25	16	0.16	78	0.78
5	2.51-3.00	2.75	8	0.08	86	0.86
6	3.01-3.50	3.25	5	0.05	91	0.91
7	3.51-4.00	3.75	5	0.05	96	0.96
8	4.01-4.50	4.25	2	0.02	98	0.98
9	4.51-5.00	4.75	1	0.01	99	0.99
10	5.01-5.50	5.25	1	0.01	100	1.00
Összesen			100	1.00		

gyak. N Gyakorisági hisztogramm

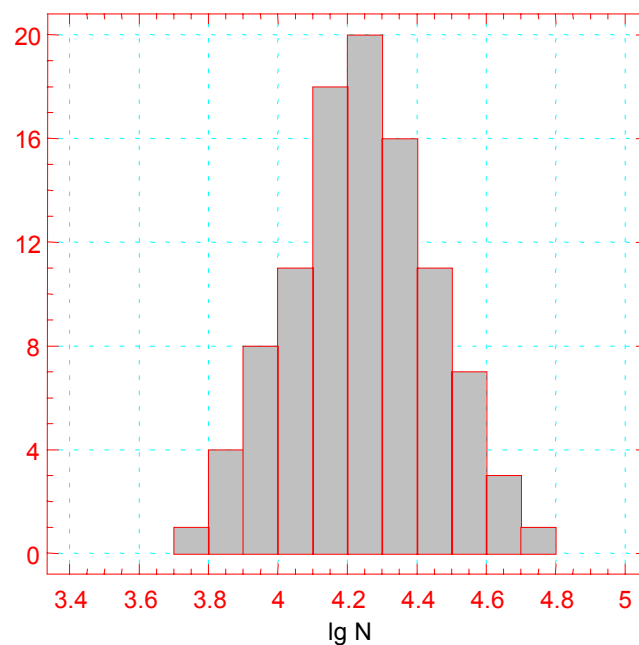


**3./3. ábra.** Homogén tejmintá sejtszám értékeinek gyakorisági hisztogramja

**3./3. Táblázat.** Mezofil aerob élősejtszám logaritmus értékeinek osztályba-sorolása

Osztály	Értékköz	Érték-közép	Gyakoriság	Relatív gyakoriság	Kumulatív gyakoriság	Kumulatív rel. gyak.
1	3,71-3,80	3,75	1	0,01	1	0,01
2	3,81-3,90	3,85	4	0,04	5	0,05
3	3,91-4,00	3,95	8	0,08	13	0,13
4	4,01-4,10	4,05	11	0,11	24	0,24
5	4,11-4,20	4,15	18	0,18	42	0,42
6	4,21-4,30	4,25	20	0,20	62	0,62
7	4,31-4,40	4,35	16	0,16	78	0,78
8	4,41-4,50	4,45	11	0,11	89	0,89
9	4,51-4,60	4,55	7	0,07	96	0,96
10	4,61-4,70	4,65	3	0,03	99	0,99
11	4,71-4,80	4,75	1	0,01	100	1,00
Összesen			100	1,00		

gyak. lg N Gyakorisági hisztogramm



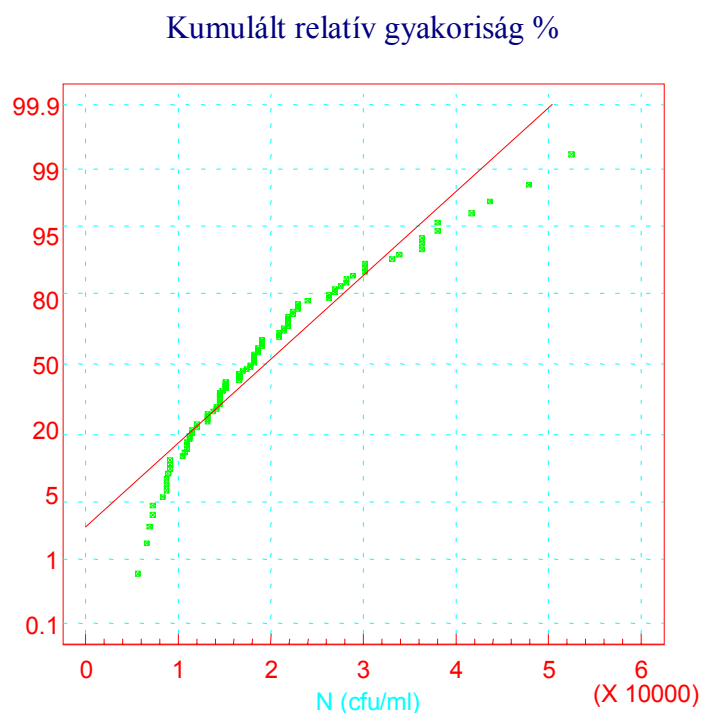
**3./4. ábra.** Homogén tejmintálg N értékeinek gyakorisági hisztogramja

A mikrobaszám alapadatok és azok logaritmikus transzformációjával nyert értékek láthatóan eltérő eloszlásúak. Az alapadatok eloszlása erősen aszimmetrikus, a logaritmus értékek eloszlása igen jó közelítéssel normálisnak látszik.

Az eloszlások normalitásvizsgálatának **nullhipotézise**, hogy a vizsgált jellemző sokaságon belüli eloszlása normális. Ez a feltétel akkor teljesül, ha a sokaságból vett mintaelemekből becsült átlag- és szórásértékek ismeretében, a mintaelemek gyakorisága a vizsgált jellemzőre a normális eloszlás 2.2./1. összefüggéssel megadott sűrűségfüggvényét követi.

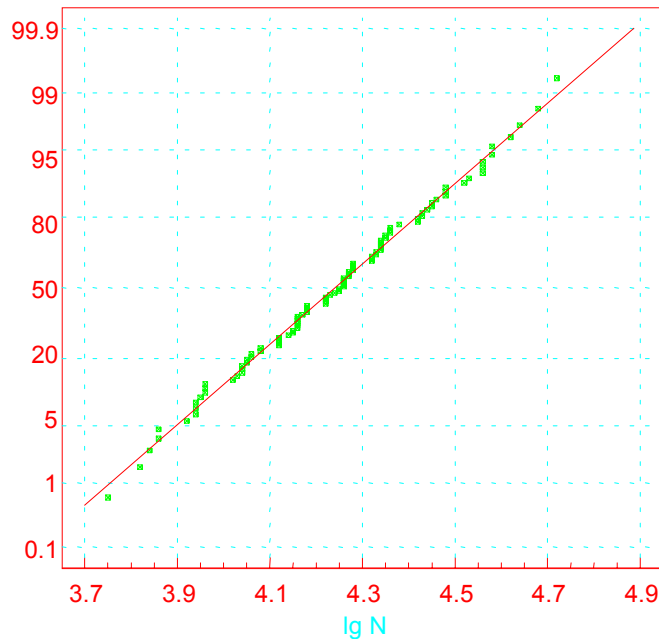
A normalitás teljesülésének grafikus szemléltetésére szolgál, ha az adatokból számított (3./2. és 3./3. táblázatokban összefoglalt) kumulatív relatív gyakoriságokat hasonlítjuk össze a normál eloszlás esetén elméletileg várható értékekkel. Amennyiben a vizsgált jellemző eloszlása normálisnak tekinthető, a kumulált relatív gyakorisági értékek a normális eloszlásra jellemző egyenes mentén oszlanak el.

A mikrobaszámokra vonatkozó összehasonlítást a 3./5. és 3./6. ábrák szemléltetik.



**3./5. ábra.** Mikrobaszám alapadatok (N) normális eloszlásának grafikus vizsgálata

### Kumulált relatív gyakoriság %



**3./6. ábra.** lg N értékek normális eloszlásának grafikus vizsgálata

A 3./5. és 3./6. ábrák vizsgálata egyértelműen szemlélteti a mikrobaszám alapadatok normalitástól való eltérését és a logaritmus transzformáció normális eloszlást eredményező hatását.

A normalitástól való eltérés matematikai-statisztikai bizonyítását  $\chi^2$ -próbával végezzük el. A próba menete a következő:

Az adatokat osztályközökbe soroljuk és meghatározzuk az egyes osztályközökbe tartozó gyakoriságokat ( $f_i$ ). A mintaelemekből számított átlag- és szórásérték ismeretében kiszámítjuk az osztályközök alsó és felső határához tartozó standardizált  $x_{ist} = (x_i - x_{\text{át}})/s$  értékeket, amelyek a standardizált normális eloszlás  $u$  értékeinek becslésére szolgálnak, majd meghatározzuk a hozzájuk tartozó  $\Phi(u)$  értéket. Az értéktartományok alsó és felső határához megállapított  $\Phi(u)$  értékek különbsége adja az elméleti relatív gyakoriságot ( $p_i$ ). A  $p_i$  értékek és a mintaelemszám szorzata megadja az egyes értékközök (osztályok) elméleti gyakoriságát ( $f_{\text{elm}}$ ). A megfigyelt és az elméleti gyakoriságok különbségéből számítható eltérésnégyzetek összege  $\chi^2$ -eloszlású. Összeadva az egyes osztályközökhöz tartozó  $\chi^2$  értékeket, kapjuk a statisztikai próba számított  $\chi^2$  értékét. Amennyiben ez nagyobb, mint az osztályköz – 3 szabadsági fokhoz tartozó kritikus (táblázatos) érték, az eloszlás az adott valószínűségi szinten szignifikánsan eltér a feltételezett normális eloszlástól.

A továbbiakban a rész-számításokat mellőzve, csak a próbák összefoglaló számítógépes táblázatait közöljük.

**3.4. Táblázat.** Mezofil aerob élősejtszám értékek (N) normalitás-vizsgálatának  $\chi^2$  próbája.

Alsó határ	Felső határ	Számlált gyakoriság	Elméleti gyakoriság	$\chi^2$
	5 000	0	6.4	6.390
5 000	10 000	13	9.6	1.217
10 000	15 000	25	16.0	5.049
15 000	20 000	24	20.4	0.644
20 000	25 000	16	19.8	0.717
25 000	30 000	8	14.6	2.990
30 000	35 000	5	8.2	1.267
35 000		9	5.0	3.118
<b>Összesen</b>				<b>21.3913</b>

Szabadsági fok = 5

Szignifikancia-szint =  $6.83 \cdot 10^{-4}$

**Döntésünk:** 0.068 % -nál nagyobb elsőfajú hibát megengedve, nullhipotézisünk már nem teljesül, azaz 99.9 %-os biztonsággal állíthatjuk, hogy az eloszlás szignifikánsan eltér a normális eloszlástól.

A 95 %-os biztonsági szinthez tartozó kritikus érték:  $\chi^2 = 11.1$

**3.5. Táblázat.** lg N értékek normalitás-vizsgálatának  $\chi^2$  próbája.

Alsó határ	Felső határ	Számlált gyakoriság	Elméleti gyakoriság	$\chi^2$
	4.0	13	12	0.0626
4.0	4.1	11	13	0.2029
4.1	4.2	18	17	0.0180
4.2	4.3	20	19	0.0367
4.3	4.4	16	17	0.0299
4.4	4.5	11	12	0.0272
4.5		11	10	0.0342
<b>Összesen</b>				<b>0.4115</b>

Szabadsági fok = 4

Szignifikancia-szint = 0.9815

**Döntésünk:** nullhipotézisünket csak 98.15 % elsőfajú hiba felett utasíthatjuk el. Annak a valószínűsége, hogy az eloszlás eltér a normálistól, kisebb, mint 1.85 %, azaz 98.1 %-os biztonsággal állíthatjuk, hogy a lg N értékek eloszlás nem tér el szignifikánsan a normálistól.

A 95 %-os biztonsági szinthez tartozó kritikus érték:  $\chi^2 = 9.49$

### 3.1.2. Várható érték becslésére szolgáló jellemzők

#### Számtani átlag ( $\bar{x}$ )

Az alapsokaság várható értékének becslésére szolgál a minta átlagértéke (tapasztalati középérték, algebrai átlag, számtani átlag). A továbbiakban mi a többi átlagértéktől való megkülönböztetés érdekében a számtani átlag kifejezést használjuk.

A **számtani átlag** a következő módon számítható:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (3.1./1.)$$

A számtani átlag a következő fontos tulajdonságokkal rendelkezik

- Minden átlagérték közül a számtani átlag a legnagyobb.
- A számtani átlagtól való eltérések összege 0, az eltérés négyzetek összege pedig minimális.

$$\sum_{i=1}^n (x_i - \bar{x}) = 0 \quad (3.1./2.)$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \min. \quad (3.1./3.)$$

Ez utóbbi megfogalmazás azt jelenti, hogy bármely más átlag esetében az eltérések négyzetének összege nagyobb, mint a számtani átlaggal számított érték.

- Tetszőleges eloszlású,  $\mu$  várható értékű és  $\sigma^2$  szórásnégyzetű alapsokaságból származó  $n$  elemű minták számtani átlaga, mint valószínűségi változó  $\mu$  várható értékű és  $\sigma^2/n$  szórásnégyzetű normális eloszlást követ. Amennyiben a mintaelemszám ( $n$ ) összemérhető az alapsokaság ( $N$ ) elemszámával, az átlagértékek szórássának pontos számítása:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}}, \quad (3.1./4.)$$

ahol  $n/N$  az un. reprezentációs arány. Ha  $n/N < 0.05$  akkor az átlagértékek szórássának számítása:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (3.1./5.)$$

#### Geometriai átlag ( $\bar{x}_g$ )

matematikai definíciója: 
$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} \quad (3.1./6.)$$

Számítása a logaritmus azonosságok felhasználásával történik:



$$\lg \bar{x}_g = \frac{1}{n} \sum_{i=1}^n \lg x_i = \frac{\lg x_1 + \lg x_2 + \dots + \lg x_n}{n} \quad (3.1./7.)$$

A (3.1./7.) összefüggésből egyértelműen kitűnik, hogy a geometriai átlag a logaritmikus átlag visszatranszformált értéke:

$$\bar{x}_g = 10^{(\lg \bar{x}_g)} \quad (3.1./8.)$$

### **Harmonikus átlag ( $\bar{x}_h$ )**

matematikai definíciója: 
$$\frac{1}{\bar{x}_h} = \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} = \frac{1}{n} \cdot \left( \frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right) \quad (3.1./9.)$$

### **Tapasztalati medián ( $Me$ )**

A medián, vagy centrális érték a nagyság szerint rendezett mintaelemek középső eleme, ha n páratlan. Ha n páros, akkor a két középső elem számtani közepe.

### **Tapasztalati módusz ( $Mo$ )**

A módusz a mintában leggyakrabban előforduló elem

Az alapsokaság várható értékének becslésére szolgáló jellemzők közül a számtani átlag a legmegbízhatóbb (torzítatlan, hatékony, konzisztens és elégséges). Minden más átlagérték, tekintettel arra, hogy kisebb, mint a számtani átlag, a várható értékre torzított becslést ad.

A medián csak szimmetrikus eloszlás esetében ad torzítatlan becslést, de ennek hatékonysága az átlagértékéhez viszonyítva csupán 63,7%-os.

A 3./1. Táblázatban összefoglalt élősejtszám adatokból (N) és a lgN értékekből kiszámított átlagokat a 3./6. Táblázatban foglaltuk össze. A táblázatban feltüntettük a lgN adatokból számított átlagok visszatranszformált értékeit (N\*) is.

**3./6. Táblázat.** A 3./1. táblázatban lévő adatok átlagértékeinek összehasonlítása.

	N	lg N	N*
Számtani átlag:	19400	4.241	17420
Medián:	17990	4.255	17990
Geometriai átlag	17406	4.235	17180
Harmonikus átlag:	15581	4.231	17020

Az alapadatok (N) geometriai átlagértékének számértéke (17406) tökéletesen megegyezik a sejtszám logaritmusok kerekítés nélküli átlagából (4.2407) visszatranszformált értékkel. Az N értékek mediánjára vonatkozóan természetesen mind az alapadatokból, mind a lgN adatokból visszatranszformálva ugyanazt az értéket kapjuk.

### 3.1.3. Szórás becslésére szolgáló jellemzők

#### *Tapasztalati szórásnégyzet ( $s_n^2$ )*

A tapasztalati szórásnégyzet a mintaelemek számtani átlaguktól való eltéréseinek számtani átlaga. Számításának menete:

$$s_n^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{1}{n} \cdot \left( \sum_{i=1}^n x_i \right)^2}{n} \quad (3.1./10.)$$

A tapasztalati szórásnégyzet négyzetgyöke a tapasztalati szórás ( $s_n$ ) azonban nem ad torzítatlan becslést az alapsokaság  $\sigma$  szórására vonatkozóan. Különösen igaz ez kis mintaszámok esetén.

#### *Korrigált tapasztalati szórásnégyzet ( $s^2$ )*

A korrigált tapasztalati szórásnégyzet torzítatlan és konzisztens becslését adja az alapsokaság  $\sigma^2$  szórásnégyzetének. Kiszámításakor az eltérésnégyzetek összegét  $(n-1)$ -gyel osztjuk:

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{1}{n} \cdot \left( \sum_{i=1}^n x_i \right)^2}{n-1} \quad (3.1./11.)$$

$$s^2 = \frac{n}{n-1} \cdot s_n^2 \quad (3.1./12.)$$

A (3.1./12.) összefüggésből jól látható, hogy a korrigált tapasztalati szórás mindig nagyobb a tapasztalati szórásnál, a különbség azonban a mintaszám növelésével csökken.

A korrigált tapasztalati szórásnégyzet a következő fontos tulajdonságokkal rendelkezik.

- A  $\mu$  várható értékű és  $\sigma^2$  szórásnégyzetű normális eloszlású alapsokaságból származó  $n$  elemű mintából számított  $\frac{(n-1) \cdot s^2}{\sigma^2}$  valószínűségi változó  $(n-1)$  szabadsági fokú  $\chi^2$  eloszlást követ.
- Az  $\frac{\bar{x} - \mu}{s} \cdot \sqrt{n}$  valószínűségi változó  $(n-1)$  szabadságfokú Student féle  $t$ -eloszlást követ.

A számtani átlag jellemző tulajdonságainál ismertettekhez hasonlóan, az átlagértékek szórása a korrigált tapasztalati szórással a következő módon becsülhető:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \quad (3.1./13.)$$

A matematikai statisztikai gyakorlatban a korrigált tapasztalati szórás ( $s$ ) standard eltérésként (Standard Deviació, SD), míg az  $s/\sqrt{n}$  kifejezés standard hiba (Standard error) néven szerepel. Mivel a statisztikai próbákban szinte kizárólag a korrigált tapasztalati szórást alkalmazzák, szórás

fogalma alatt általánosságban ezt értjük. A továbbiakban az egyszerűség kedvéért, ha csak külön nem jelezzük, a mintára vonatkozóan a korrigált tapasztalati szórásra a szórás kifejezést alkalmazzuk

### **Relatív szórás**

A relatív szórás, vagy variációs együttható (CV) a szórás és az átlagérték hányadosa ( $s/\bar{x}$ ), amelyet általában százalékosan adnak meg.

A 3./1. táblázatban szereplő élősejtszám adatok átlagértékeit, szórásait és relatív szórásait a 3./7. táblázatban foglaltuk össze.

### **3./7. Táblázat.** A 3./1 táblázatban összefoglalt adatok szórás értékei

	N	lg N
Átlagérték	19400	4.241
Szórás (SD)	9481	0.206
Relatív szórás %	48.8	4.86

A lgN adatokból számított szórás a számításnál felhasznált logaritmikus azonosságok következtében tulajdonképpen az alapadatok relatív szórására enged következtetni. Esetünkben ez az érték:  $10^{0.206} = 1.607$ , ami a logaritmikus átlagból visszatranszformált értékhez viszonyítva  $\pm 60.7$  %-os eltérést jelent. Ez jelentősen nagyobb, mint a normál alapadatok relatív szórása (48.8 %). Az eltérés oka, hogy a logaritmikus átlagból visszatranszformált érték (az alapadatok geometriai átlaga) mindig kisebb, mint a számtani átlag.

### **Mintaterjedelem (R)**

A minta legnagyobb és legkisebb elemének különbsége. A mintaterjedelem lehetőséget ad az alapsokaság szórásának közelítő becslésére. A becslés torzított és hatékonysága (E) a mintaszám növelésével romlik. A mintaterjedelem alapján történő becslés nagy információvesztéssel jár, de gyors tájékozódást tesz lehetővé.

A szórás a mintaterjedelemből egy szorzófaktor segítségével becsülhető:  $s = fR$ , ahol f értéke függ a mintaelemek (n) számától. Az összefüggés jellemzőit a 3./8. táblázatban foglaltuk össze.

### **3./8. Táblázat.** Szórás becslése mintaterjedelem alapján

n	f	E
2	0.886	1.00
4	0.486	0.95
6	0.395	0.93
8	0.351	0.89
10	0.325	0.85
20	0.268	0.70

### 3.2. AZ ALAPSOKASÁG PARAMÉTEREINEK BECSLÉSE

Az alapsokaság paramétereinek becslésére a mintából meghatározott statisztikai jellemzők szolgálnak, melyeket a fentiekben ismertettünk. Az alábbiakban csak a legfontosabb, témakörünket érintő eloszlástípusok paramétereinek becslésére szolgáló statisztikai jellemzőket foglaljuk össze.

#### 3.2.1. Normális eloszlású alapsokaság paramétereinek becslése

*Várható érték ( $\mu$ ) becslése*

A *számtani átlag* ( $\bar{x}$ ) torzítatlan, hatékony, konzisztens és elégséges becslése az alapsokaság várható értékének.

A *medián* szintén konzisztens és növekvő  $n$  esetén torzítatlan becslése a várható értéknek, de nagyobb a szórása, mint a számtani középértéknek.

*Szórásnégyzet ( $\sigma^2$ ) becslése*

A *korrigált tapasztalati szórásnégyzet* ( $s^2$ ) torzítatlan és konzisztens becslést ad.

A tapasztalati szórásnégyzet ( $s_n^2$ ) és a terjedelem ( $R$ ) torzított becslést ad.

#### 3.2.2. Binomiális eloszlás paramétereinek becslése

*A valószínűség ( $p$ ) becslése*

A *relatív gyakoriság* torzítatlan, konzisztens és elégséges becslést ad:  $p \approx k/n$ .

*A szórásnégyzet becslése*

A becslés szórásnégyzete:  $\sigma^2 \approx n \cdot p \cdot (1-p)$

#### 3.2.3. Poisson eloszlás $\lambda$ paraméterének becslése

A Poisson eloszlás várható értéke és szórásnégyzete is  $\lambda$ , ezért ennek becslésére a minta számtani átlaga és korrigált tapasztalati szórásnégyzete egyaránt használható. Mindkét becslés torzítatlan és konzisztens, de a számtani közép hatékonysága jobb, mert ennek szórása kisebb.

### 3.3. KONFIDENCIA-INTERVALLUM SZÁMÍTÁSOK

Az alapsokaság ismeretlen, (de konkrét  $a$  értékű) paraméterét a mérési eredmények valószínűségi változó jellege miatt nem tudjuk pontosan meghatározni. A mérési eredmények statisztikai jellemzői (átlag, szórás) alapján azonban lehetőségünk van megadni egy olyan tartományt, *konfidencia-intervallumot*, amelyen belül az alapsokaság ismeretlen  $a$  paramétere egy általunk meghatározott valószínűséggel (megbízhatósággal) elhelyezkedik. Természetesen minél nagyobb valószínűséggel (megbízhatósággal) kívánjuk megadni ezt a tartományt, annál szélesebbre kell választani.

Olyan ( $a_1$ ;  $a_2$ ) intervallumot keresünk, amelyen belül az  $a$  paraméter  $1-\alpha$  valószínűséggel megtalálható:

$$P(a_1 \leq a < a_2) = 1-\alpha,$$

ahol  $\alpha$  az általunk megengedett elsőfajú hiba nagysága (a statisztikai gyakorlatban általában 5%, ritkábban 1, ill. 0.1%). Az  $a_1$  és  $a_2$  értékek valószínűségi változók, melyeknek konkrét értékét a mintából számított statisztikai jellemzők segítségével határozzuk meg. Az ( $a_1$ ;  $a_2$ ) intervallumot az  $a$  paraméterre vonatkozó  $1-\alpha$  megbízhatósági szintű *konfidencia-intervallumnak* nevezzük.

### 3.3.1. Normál eloszlással kapcsolatos konfidencia-intervallum számítások

#### 3.3.1.1. Várható érték konfidencia-intervaluma ismert szórás esetén

Mint azt a 3.1.2. pontban ismertettük, a minta számtani átlaga, mint valószínűségi változó,  $\mu$  várható értékű és  $\sigma^2/n$  szórásnégyzetű normális eloszlást követ. Ennek megfelelően a belőle képzett

$$u = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} \quad (3.3./1.)$$

valószínűségi változó standardizált normális eloszlású, melynek várható értéke 0, szórása 1.

Kiválasztva a megfelelő valószínűségi szintet, a standardizált normális eloszlás eloszlásfüggvény táblázatából, attól függően, hogy egyoldali, vagy kétoldali elsőfajú hibát engedünk meg  $u_{1-\alpha}$ , vagy  $u_{1-\alpha/2}$  kikereshető.

Leggyakoribb esetben, 5 %-os elsőfajú hibát választva,  $\alpha = 0.05$ , ill.  $\alpha/2 = 0.025$ . Az eloszlás kétoldalán megengedve az  $\alpha/2$  elsőfajú hibát (ld. 3./1. ábra), a standardizált normáloszlás eloszlásfüggvényének táblázatában  $\Phi(u)$  értéke 0,975 (=1-0,025). a hozzátartozó  $u$  érték: 1,96.

Az  $u$  értékének ismeretében a (3.3./1.) összefüggésből a  $\Delta = |\bar{x} - \mu|$  eltérés kiszámítható.

$$\Delta = u \cdot \frac{\sigma}{\sqrt{n}} \quad (3.3./2.)$$

Kétoldali eltérést megengedve, az alapsokaság  $\mu$  várható értéke  $1-\alpha$  valószínűséggel az  $\bar{x} \pm \Delta$  intervallumban lesz. Részletesen kifejtve, a  $\mu$  várható érték konfidencia-intervalluma az alábbi módon számolható:

$$\bar{x} - u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu < \bar{x} + u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \quad (3.3./3.)$$

ahol  $\frac{\sigma}{\sqrt{n}}$  az átlagértékek szórása.

A (3.3./3.) összefüggésből jól látható, hogy a konfidenciaintervallum nagysága az alapsokaság adott szórástól és a mintaszám nagyságától függ.

A (3.3./2.) összefüggés lehetőséget nyújt arra, hogy kiszámíthassuk a várható érték  $\Delta$  pontosságú becsléséhez szükséges mintaelemek számát:

$$n = \left( u \cdot \frac{\sigma}{\Delta} \right)^2 \quad (3.3./4.)$$

#### 3.3.1.2. Szórás és várható érték konfidencia-intervaluma ismeretlen szórás esetén

### **Alapsokaság szórásának konfidencia-intervalluma**

A  $\mu$  várható értékű és  $\sigma^2$  szórásnégyzetű normális eloszlású alapsokaságból származó  $n$  elemű mintából számított  $\frac{(n-1) \cdot s^2}{\sigma^2}$  valószínűségi változó  $(n-1)$  szabadsági fokú  $\chi^2$  eloszlást követ (3.1.3. pont), ami lehetőséget teremt  $\sigma^2$  konfidencia-intervallumának a mintából számított szórásnégyzet ( $s^2$ ) alapján való meghatározására.

$$\frac{n-1}{\chi_{1-\alpha/2}^2} \cdot s^2 < \sigma^2 \leq \frac{n-1}{\chi_{\alpha/2}^2} \cdot s^2 \quad (3.3./5.)$$

Ha a mintaszám ( $n$ ) nagyobb, mint 100, akkor a minta szórásának szórása:  $s_s = s/\sqrt{2n}$  és a szórás eloszlása normálishoz közelít. Ebben az esetben az alapsokaság szórásának konfidencia-intervalluma a standardizált normális eloszlás felhasználásával a (3.3./3.) összefüggés analógiájára:

$$s - u_{1-\alpha/2} \cdot \frac{s}{\sqrt{2n}} < \sigma \leq s + u_{1-\alpha/2} \cdot \frac{s}{\sqrt{2n}} \quad (3.3./6.)$$

### **Alapsokaság várható értékének konfidencia-intervalluma**

Kihasználva, hogy az  $\frac{\bar{x} - \mu}{s} \cdot \sqrt{n}$  valószínűségi változó  $(n-1)$  szabadságfokú Student féle  $t$ -eloszlást követ, a 3.3.1.1. pontban ismertetett eljárással analóg módon végezzük a számításokat. Az eltérés csupán annyi, hogy a (3.3./3.) összefüggésben  $u$  helyett a Student féle  $t$ -eloszlás  $t$  értékét, valamint  $\sigma$  helyett a mintából számított  $s$  értéket használjuk a konfidencia-intervallum kiszámítására:

$$\bar{x} - t \cdot \frac{s}{\sqrt{n}} \leq \mu < \bar{x} + t \cdot \frac{s}{\sqrt{n}} \quad (3.3./7.)$$

Hasonló módon a várható érték  $\Delta$  pontosságú kiszámításához szükséges mintaszám:

$$n = \left( t \cdot \frac{s}{\Delta} \right)^2 \quad (3.3./8.)$$

### **Példák konfidencia-intervallum számításra**

A konfidencia-intervallumok számítását a 3.1. táblázatban feltüntetett élősejtszám adatok kiértékelésén keresztül mutatjuk be. A mintából számított jellemzőket a 3.9. táblázatban foglaltuk össze.

**3./9. Táblázat.** A 3./1 táblázatban összefoglalt adatok statisztikai jellemzői

	N	lg N
Átlagérték	19400	4.241
Szórás (SD)	9481	0.206
Minta-elemszám	100	100

**Példa a szórás konfidencia-intervallumának számítására**

A 3./1. táblázatban összefoglalt adatok  $n=100$  mintaszáma lehetővé teszi, hogy a normális eloszlású alapsokaság szórásának konfidencia-intervallumát mindkét módon (standardizált normál eloszlás és  $\chi^2$  eloszlás felhasználásával) egyaránt kiszámítsuk. Mivel a sejttség (N) értékek eloszlása bizonyítottan nem normális (ld. 3.1.1. pont), a szórás konfidencia-intervallumát csak a normális eloszlású lg N értékekre vonatkozóan határozzuk meg.

Szabadsági fok =  $n-1 = 99$ . Választott megbízhatósági szint 95 %.

Kétoldali  $\chi^2$  értékek:  $\chi_{97,5}^2 = 74$       $\chi_{2,5}^2 = 129$ .      $u = 1,96$

lg N adatokból számított értékek

$$\chi^2 \text{ eloszlás alapján számítva:} \quad \frac{99}{129} \cdot 0.206^2 < \sigma^2 \leq \frac{99}{74} \cdot 0.206^2$$

$$\text{A } \sigma_{\lg N} \text{ érték 95 \% -os konfidencia-intervalluma:} \quad 0.1806 < \sigma_{\lg N} \leq 0.2383$$

$$\text{Standardizált normál eloszlás alapján:} \quad 0,206 - 1.96 \cdot \frac{0.206}{\sqrt{200}} < \sigma_{\lg N} \leq 0,206 + 1.96 \cdot \frac{0.206}{\sqrt{200}}$$

$$\text{A } \sigma_{\lg N} \text{ érték 95 \% -os konfidencia-intervalluma:} \quad 0.1774 < \sigma_{\lg N} \leq 0.2346$$

Összehasonlítva a korrekt kiértékeléssel ( $\chi^2$  eloszlás) kapott konfidencia-határokat a közelítő eredményt biztosító (standardizált normál eloszlás) számítással kapott értékekkel, megállapítható, hogy a közelítő, de jelentősen egyszerűbb számítással kapott konfidencia-határok mintegy 1,5 %-kal kisebbek. Az eltérés a  $\chi^2$  eloszlásnak az adott szabadsági foknál még tapasztalható aszimmetriájából fakad.

#### **Példa a várható érték konfidencia-intervallumának számítására**

A várható érték konfidencia-intervallumát ismert szórás esetén a (3.3./3.) összefüggéssel, ismeretlen szórás esetén pedig a (3.3./7.) összefüggéssel számítjuk ki. Példánkban ez utóbbi esettel foglalkozunk a 3.9. táblázat adatainak felhasználásával.

A Student féle t-eloszlás határesetben a standardizált normáeloszlásba megy át. Végtelen szabadsági foknál (de gyakorlatilag már 120 felett) a t értékek helyett az azonos valószínűségi szinthez tartozó u értékek használhatók. Példánk esetében a 99-es szabadsági foknál a 95 %-os biztonsági szinthez (mindkét oldalon 2,5 % elsőfajú hibával) tartozó t érték 1.99, jó közelítését adja az 1,96-os u értéknek.

A centrális határeloszlás tétele következtében a mintaátlagok akkor is normális eloszlást követnek, ha az alapsokaság nem normális eloszlású, ezért a várható érték konfidencia-intervalluma mind a sejtstszámokra (N), mind pedig azok logaritmusára (lg N) kiszámítható.

Behelyettesítve a megfelelő adatokat a (3.3.7.) összefüggésbe, a várható érték konfidencia-intervalluma az alábbiak szerint számítható.

### Alapadatokból számított értékek

$$19400 - 1.99 \cdot \frac{9481}{\sqrt{100}} \leq \mu_N < 19400 + 1.99 \cdot \frac{9481}{\sqrt{100}}$$

A  $\mu_N$  érték 95 %-os konfidencia-intervalluma:  $17513 \leq \mu_N < 21287$

### lg N adatokból számított értékek

$$4.241 - 1.99 \cdot \frac{0.206}{\sqrt{100}} \leq \mu_{\lg N} < 4.241 + 1.99 \cdot \frac{0.206}{\sqrt{100}}$$

A  $\mu_{\lg N}$  érték 95 %-os konfidencia-intervalluma:  $4.200 \leq \mu_{\lg N} < 4.282$

### ***Példa a várható érték adott pontosságú becsléséhez szükséges mintaszám számítására***

A várható érték  $\Delta$  intervallumú pontossággal való kiszámításához szükséges mintaszám a (3.3./8.) összefüggés szerint:  $n = \left( t \cdot \frac{s}{\Delta} \right)^2$ , ahol  $t$  értékének szabadsági foka mindenkor megegyezik a szórás számításához felhasznált szabadsági fokkal.

A  $\Delta$  értékének dimenziója megegyezik a várható érték dimenziójával, amit a megkívánt pontosság előírásakor feltétlenül figyelembe kell venni.

Ha relatív eltérést (pl. a várható érték 20 %-át írjuk elő konfidencia-intervallumként, ami megfelel  $\pm 10$  % eltérésnek), akkor a mikrobaszám alapadatok esetében ehhez szükségünk van a várható érték becslésére is. Logaritmált adatok relatív eltérése a várható érték ismerete nélkül is meghatározható a logaritmikus azonosságok felhasználása révén. (Pl. 10 %-os relatív eltérés megfelel 0,0414 logaritmus egységnek, mert  $\lg 1,10 = 0,0414$ .)

Számítsuk ki a továbbiakban, hogy a 3./9. táblázatban összefoglalt jellemzők alapján hány minta szükséges a várható érték  $\pm 10$  %-os konfidencia-intervallumának meghatározásához.

### Normál adatok (N) felhasználásával számolva

Az átlagérték 10 %-a  $\Delta_N = 1940$ , behelyettesítve az adatokat:

$$n = \left( 1.99 \cdot \frac{9481}{1940} \right)^2 = 94,6, \text{ azaz a szükséges minta elemszám: } n \geq 95.$$

### Logaritmált (lg N) adatok felhasználásával számolva

Az átlagérték 10 %-a megfelel 1,10-zel való szorzásnak/osztásnak, ami  $\pm 0,0414$  logaritmus egységet jelent. Ennek megfelelően  $\Delta_{\lg N} = 0,0414$ . Behelyettesítve az adatokat:

$$n = \left( 1.99 \cdot \frac{0.206}{0.0414} \right)^2 = 98,0, \text{ azaz a szükséges minta elemszám: } n \geq 98.$$

A két mintaszám közel megegyezik egymással.



### 3.3.2. Binomiális eloszlással kapcsolatos konfidencia-intervallum számítások

Valamely esemény  $n$  számú megfigyelésből  $k$  alkalommal való bekövetkezése esetén ( $p_1$   $n$  db konzervből  $k$  db. fertőzött) az előfordulás valószínűségének becslésére a  $p = k/n$  hányados szolgál. Tekintettel arra, hogy a  $k$  valószínűségi változó diszkrét érték, a felhasználásával számított és a  $p$  valószínűség becslésére szolgáló  $k/n$  hányados sem lehet folytonos változó, a konfidencia-intervallumot úgy határozzuk meg, hogy az legalább  $1-\alpha$  szintű legyen, azaz

$$P(p_1 \leq p < p_2) \geq 1 - \alpha$$

A  $p_1$  és  $p_2$  konfidencia-határokat megadó összefüggések:

$$\sum_{r=k}^n \binom{n}{r} p_1^r (1-p_1)^{n-r} = \frac{\alpha}{2} \quad (3.3./9.)$$

$$\sum_{r=0}^k \binom{n}{r} p_2^r (1-p_2)^{n-r} = \frac{\alpha}{2} \quad (3.3./10.)$$

A két összefüggés egy-egy egyenletet szolgáltat  $p_1$  és  $p_2$  értékének meghatározására, amely megoldásokat táblázatos formában szokás megadni. Mivel a binomiális eloszlás minden  $p$  és  $n$  értékre más, a binomiális eloszlásokat összefoglaló táblázatok igen terjedelmesek. Gyakorlati szempontból azonban elég csupán a  $k = 0 - 20$  és  $n-k = 0 - 20$  értékekig megadni a  $p_1$  és  $p_2$  konfidencia-határokat, mert nagy megfigyelésszám esetén, a binomiális eloszlás jól közelíthető a  $\mu = np$  várható értékű és  $D = \sqrt{n \cdot p \cdot (1-p)}$  szórású normális eloszlással (ld. 2.1.2. fejezet).

A  $\frac{k - np}{\sqrt{np(1-p)}}$  valószínűségi változó közel normális eloszlású,  $\mu = 0$  várható értékkel és  $\sigma = 1$  szórással, ennek megfelelően  $p_1$  és  $p_2$  kiszámítására a standardizált normál eloszlás használható.

$$P\left(-u_{1-\alpha/2} \leq \frac{k - np}{\sqrt{np(1-p)}} \leq u_{1-\alpha/2}\right) \approx 1 - \alpha \quad (3.3./11.)$$

$u_{1-\alpha/2} - t$  u-val jelölve és megoldva a (3.3./11.) egyenletet  $p_1$  és  $p_2$ -re:

$$p_1 = \frac{\frac{k}{n} + \frac{u^2}{2n} - \frac{u}{\sqrt{n}} \sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right) + \frac{u^2}{4n}}}{1 + \frac{u^2}{n}} \quad (3.3./12.)$$

$$p_2 = \frac{\frac{k}{n} + \frac{u^2}{2n} + \frac{u}{\sqrt{n}} \sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right) + \frac{u^2}{4n}}}{1 + \frac{u^2}{n}} \quad (3.3./13.)$$

### ***Példa binomiális eloszlás konfidencia-intervallumára.***

Konzervek termosztátpróbáját végezve,  $n=20$  mintából  $k=5$  esetben tapasztaltak romlást. Az ennek megfelelő romlási arány:  $p=k/n=0,25$ . Határozzuk meg a várható romlási arány alapsokaságon belüli 95 %-os konfidencia-intervallumát.

A binomiális eloszlás táblázatából  $k=5$  és  $n-k=15$  adatpárokhoz tartozóan a  $p$  valószínűség 95%-os konfidencia-határai:  $p_1 = 0,087$  és  $p_2 = 0,491$ , azaz valódi romlási arány:  $8,7\% \leq p < 49,1\%$  intervallumban található, ami meglehetősen széles sáv.

A selejtarány konfidencia-intervallumának szűkítése érdekében ismételt mintavételt végezve,  $n=200$  db konzervből a termosztátpróba során  $k=40$  bizonyult hibásnak. Az ennek megfelelő romlási arány  $k/n=0,20$ . Kérdés, hogy az ismételt mintavétel alapján mekkora a tételen belüli romlási arány 95%-os konfidencia-intervalluma.

A nagy mintaszámok következtében a konfidencia-határok számítását a (3.3.12.) és (3.3./13.) összefüggésekkel határozzuk meg. A 95%-os, kétoldali valószínűségi szintnek megfelelően  $u$  értékét 1.96-nak választva:

$$p = \frac{\frac{k}{n} + \frac{u^2}{2n}}{1 + \frac{u^2}{n}} \pm \frac{\frac{u}{\sqrt{n}} \sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right) + \frac{u^2}{4n}}}{1 + \frac{u^2}{n}} \quad (3.3./14.)$$

Behelyettesítve az értékeket a (3.3./14.) egyenletbe, a következő eredményeket kapjuk.

$$p_1 = 0,1599$$

$$p_2 = 0,2703$$

azaz a tételen belüli romlási arány 95%-os konfidencia-intervalluma:  $16\% \leq p < 27\%$ .

### **3.3.3. Poisson eloszlással kapcsolatos konfidencia-intervallum számítások**

Poisson eloszlású valószínűségi változók konfidencia-intervallumának meghatározása, hasonlóan a binomiális eloszláshoz kétféle módon lehetséges, a  $\lambda$  várható értékének nagyságától függően.

Kis mintaszámok esetén a konfidencia-határok a Poisson eloszlás valószínűségi függvénye segítségével számíthatók az alábbi egyenletek megoldásaként.

$$e^{-\lambda_1} \cdot \sum_{r=k}^{\infty} \frac{\lambda_1^r}{r!} = \frac{\alpha}{2} \quad (3.3./15.)$$

$$e^{-\lambda_2} \cdot \sum_{r=0}^k \frac{\lambda_2^r}{r!} = \frac{\alpha}{2} \quad (3.3./16.)$$

Az egyenletek megoldásai a Poisson eloszlás táblázataiban találhatóak. Az  $np$ -re adott konfidencia-intervallum ebben az esetben:  $\lambda_1 \leq \lambda < \lambda_2$ .

Nagyobb  $\lambda$  értékekre vonatkozóan,  $\lambda = k/n > 15$  értékektől a Poisson eloszlás már jól közelíthető egy  $\mu = \lambda$  várható értékű és  $D^2 = \lambda$  szórásnégyzetű normális eloszlással.

A  $\frac{k - np}{\sqrt{np(1-p)}}$  valószínűségi változó közel normális eloszlású,  $\mu = 0$  várható értékkel és  $\sigma = 1$  szórással, ennek megfelelően  $p_1$  és  $p_2$  kiszámítására a standardizált normál eloszlás használható.

$$P\left(-u_{1-\alpha/2} \leq \frac{k - np}{\sqrt{np(1-p)}} \leq u_{1-\alpha/2}\right) \approx 1 - \alpha \quad (3.3./17.)$$

$u_{1-\alpha/2}$  – t  $u$ -val jelölve és megoldva a (3.3./17.) egyenletet  $p_1$  és  $p_2$ -re:

$$\lambda_1 = np_1 = \frac{k + \frac{u^2}{2} - u \cdot \sqrt{k \left(1 - \frac{k}{n}\right) + \frac{u^2}{4}}}{1 + \frac{u^2}{n}} \quad (3.3./18.)$$

$$\lambda_2 = np_2 = \frac{k + \frac{u^2}{2} + u \cdot \sqrt{k \left(1 - \frac{k}{n}\right) + \frac{u^2}{4}}}{1 + \frac{u^2}{n}} \quad (3.3./19.)$$

### ***Példa Poisson eloszlás konfidencia-intervallumára.***

Bürker kamrás sejtszámlálást végezve  $n = 5$  kamrában leszámolt élesztő sejtek száma összesen  $k = 25$ . Az egyes kamrákban lévő sejtek száma Poisson eloszlást követ, melynek várható értéke ( $\lambda$ ), amelyet az észlelések összes számával ( $k = n \cdot \lambda$ ) becsülünk.

A Poisson eloszlás táblázatából meghatározva, 95 %-os biztonsággal a  $k=25$  sejtszám szuszpenzió belüli konfidencia határai:  $k_1 = 16,2$  és  $k_2 = 36,8$ . Ezen adatokból kiszámítva a szuszpenzió sejtkoncentrációjának konfidencia-határait:

$$\lambda_1 = k_1/n = 16,2/5 = 3,24 \text{ sejt/kamra}$$

$$\lambda_2 = k_2/n = 36,8/5 = 7,36 \text{ sejt/kamra}$$

A szuszpenzió kamránkénti sejtszáma a fentiek szerint a  $3,24 \leq N < 7,36$  tartományba esik 95%-os biztonsággal. A sejtszuszenzió milliliterenkénti sejtszámát ebből az adatból a hígítás és a kamratérfogat ismeretében lehet kiszámítani.

## 4. HIPOTÉZISVIZSGÁLATOK. STATISZTIKAI PRÓBÁK

A mérési eredmények matematikai statisztikai kiértékelésekor a legtöbb esetben már van valamilyen előzetes feltevésünk (*nullhipotézis*,  $H_0$ ), amelynek helyességét, vagy helytelenségét *statisztikai próbák* elvégzésével ellenőrizzük, s döntéseinket ennek alapján hozzuk meg. Döntéseink a megfigyelt jelenségek statisztikai jellege miatt bizonytalanságot (első- és másodfajú hibát) hordoznak magukban, melynek jellegét és mértékét a statisztikai becslésekkel foglalkozó 3. fejezetben részletesen tárgyaltuk.

Az elvégzett statisztikai próbák alapján nullhipotézisünket vagy elfogadjuk, vagy elutasítjuk. Ez azonban nem jelenti azt, hogy nullhipotézisünk biztosan igaz vagy hamis. Csupán azt állíthatjuk, hogy *az adott valószínűségi szinten* a statisztikai próba eredménye nem mond ellent (vagy ellentmond) feltevésünknek. Bonyolítja a helyzetet, hogy a hipotézisvizsgálat eredménye függ az alkalmazott statisztikai próbától is. (Ezt szemléletesen bizonyítani fogjuk a módszer-összehasonlító vizsgálatok kétféle kiértékelésén keresztül.) A statisztikai próba helyes megválasztása nem csupán matematikai, hanem igen jelentős mértékben szakmai kérdés is.

A hipotézis vizsgálat általános lépései a következők.

- A mintaelemek (mérések) eredményeiből kiszámítunk egy olyan jellemzőt (*próbastatisztika*), amelynek eloszlása ismert. Ez általában a statisztikai táblázatokban is megtalálható  $u$ ,  $t$ ,  $F$ ,  $\chi^2$  stb. érték
- Meghatározzuk a döntésünknel megengedhető elsőfajú hiba mértékét. Ez általában 5, vagy 1% szokott lenni. Az elsőfajú hiba mértékének megválasztása mindig szakmai megfontolásokon alapszik. Általános irányelvként elfogadható, hogy biztonságunk növelése érdekében mindig magunk ellen döntsünk. Például ha a nullhipotézis (nincs szignifikáns különbség) teljesülése a kedvező, akkor az elsőfajú hibát válasszuk nagyra, ellenkező esetben kicsire.
- A választott elsőfajú hibához táblázatból kikereshető a próbastatisztika ( $u$ ,  $t$ ,  $F$ ,  $\chi^2$  stb.) elméleti értéke.
- A számított és a táblázatos (elméleti) értékek összehasonlítása alapján döntünk a nullhipotézisről. Amennyiben a számított érték kisebb az elméleti értéknél, a nullhipotézist elfogadjuk, s azt mondjuk, hogy a minta és a feltételezett alapsokaság eltérése nem szignifikáns. Amennyiben a számított próbastatisztika nagyobb a táblázatos értéknél, a nullhipotézist elutasítjuk, és azt mondjuk, hogy a minta eltérése a feltételezett alapsokaságtól a választott valószínűségi szinten szignifikáns.
- A számítógépes matematikai-statisztikai kiértékeléseknél a programok általában nem egy adott szignifikancia-szinthez (pl.  $\alpha=0.05$ ) adják meg a döntést, hanem a mintából számított próbastatisztikához határozzák meg a kritikus szignifikancia-szintet. Például egy statisztikai kiértékelésnél megadott  $\text{sign. level} = 0.0065$  azt jelenti, hogy a különbség minden  $\alpha > 0.0065$  esetben, vagyis 0.65% felett szignifikánsnak tekintendő.

**Kísérlettervezési szempontok.** A kísérletek megtervezésekor a felesleges munka elkerülése és a hatékony kiértékelés érdekében célszerű az alábbi sorrendet betartani.

- A vizsgált jelenségre vonatkozó nullhipotézis felállítása.
- A nullhipotézis helyességét eldöntő statisztikai próbák kiválasztása.
- A statisztikai próba hatékonyságát biztosító optimális kísérleti elrendezés kialakítása
- Kísérletek elvégzése.
- Kísérleti eredmények kiértékelése.

### 4.1. EGYMINTÁS STATISZTIKAI PRÓBÁK

Azokban az esetekben, amikor egy minta elemeiből kiszámítunk valamely paramétert és ezt hasonlítjuk össze egy elméletileg várható, vagy előírt értékkel, egymintás statisztikai próbákról beszélünk. Leggyakoribb kérdésfeltevés, hogy a mintázott sokaság várható értéke, vagy szórása megfelel-e az előírásnak.

Az egymintás statisztikai próbák szoros összefüggésben állnak a konfidencia-intervallum számításokkal. Különösen szembeűnő ez, amikor nullhipotézisünk szerint a mintából számított jellemző és az elméleti érték között nincs szignifikáns különbség. Nyilvánvaló, hogy a kérdés a vizsgált jellemző mintából számított értékének konfidencia-intervalluma alapján eldönthető. Ha az elméleti érték a számított konfidencia-intervallumon belül van, nullhipotézisünk teljesül, ellenkező esetben nem.

#### 4.1.1. Szórás összehasonlítása elméleti értékkel

A korrigált tapasztalati szórás elméleti értékhez való viszonyítása  $\chi^2$  eloszlás alapján történik.

**Nullhipotézis:** A minta az elméleti szórású sokaságból származik.

**Alternatív hipotézis:** A minta az elméletinél nagyobb szórású sokaságból származik.

**Próbastatisztika:**  $\chi^2 = S^2 \cdot (n-1) / \sigma^2$

**Összehasonlító érték:** az n-1 szabadsági fokú  $\chi^2$  érték

#### 4.1. Példa

Pasztörtej mezofil aerob élősejtszámának szórása (lg N értékekből) 100 mintaelemből számítva:  $S = 0.206$ . Előzetes felmérések alapján a gyártósorra jellemző szórás  $\sigma = 0.190$ . Kérdés: megváltozott-e a szórás?

**Nullhipotézis:** A minta az elméleti szórású sokaságból származik.

**Alternatív hipotézis:** A minta az elméletinél nagyobb szórású sokaságból származik.

**Próbastatisztika:**  $\chi^2 = 0.206^2 \cdot 99 / 0.19^2 = 116$

**Összehasonlító érték:**  $\alpha = 5\%$ , egyoldalú, Szf = 99,  $\chi^2 = 123$

**Döntés:** A számított  $\chi^2$  érték kisebb, mint a táblázatos érték, ezért nullhipotézisünket elfogadjuk. A szórás növekedése nem szignifikáns.

Fenti példánk adatbázisa azonos a szórás konfidencia-intervallumának számítására bemutatott példával (3.3.1.2. fejezet), ezért az eredmény ellenőrizhető. A 0.206 lg egységnyi szórás 95%-os konfidencia-határai: 0.1774 – 0.2346. Tekintettel arra, hogy az előírt elméleti érték (0.190) ezen határokon belül van, nyilvánvaló, hogy a szórás növekedése nem szignifikáns.

#### 4.1.2. Poisson eloszlású sokaság várható értékének statisztikai próbái

Poisson eloszlású sokaságokra vonatkozó statisztikai próbáknál kis  $\lambda$  értékeknél a Poisson eloszlás táblázati értékeit használjuk, míg  $\lambda > 15$  várható érték esetében, kihasználva, hogy ekkor az eloszlás jól közelíthető egy normális eloszlással, a standardizált normális eloszlás táblázataival dolgozhatunk.

#### 4.2. Példa kis várható értékű Poisson-eloszlásra

Csokoládé gyártósor felületi szennyezettségére vonatkozó előírás szerint a  $100 \text{ cm}^2$ -en kimutatható penész- és élesztőgomba szám nem haladhatja meg a 20-at. A felület ellenőrzése érintéses módszerrel,  $10 \text{ cm}^2$ -es agar-lemezekkel történt 3 párhuzamosban. A lemezeken kinövő telepek száma 3, 7, 5. Kérdés, hogy a felület tisztasága megfelel-e az előírásoknak, azaz az átlagos telepszám és az előírás közötti különbség szignifikáns-e.

Előírt érték: maximum 2 telep/ $10 \text{ cm}^2$ .

A 3 lemez átlagértéke:  $15/3 = 5$  telep/ $10 \text{ cm}^2$ .

Poisson eloszlású gyakoriságok táblázatából 5%-os elsőfajú hibát választva, a  $k = \sum k_i = 15$  értékhez tartozó konfidencia-határok 8.4 és 24.8. Ebből számítva az átlagértékek konfidencia-határai:  $\lambda_1 = 8.4/3 = 2.8$ ,  $\lambda_2 = 24.8/3 = 8.3$

Tekintettel arra, hogy a mintából számított felületi szennyezettség alsó konfidencia-határa magasabb, mint az előírt érték, a szennyezettség szignifikánsan nagyobb az előírtnál.

#### 4.3. Példa nagy várható értékű Poisson-eloszlásra

Az előző példában szereplő üzem felületi higiéniai előírása mezofil aerob mikrobaszámra vonatkozóan  $100 \text{ cm}^2$ -en maximum 200 mikrobát engedélyez. A felület ellenőrzése érintéses módszerrel,  $10 \text{ cm}^2$ -es agar-lemezekkel történt 3 párhuzamosban. A lemezeken kinövő telepek száma: 25, 50, 38.

Előírt érték: maximum 20 telep/ $10 \text{ cm}^2$ .

A 3 lemez átlagértéke:  $113/3 = 37.7$  telep/ $10 \text{ cm}^2$ .

Poisson-eloszlás  $\lambda > 15$  várható értékeknél jól közelíthető egy olyan normális eloszlással, melynek várható értéke  $\mu = \lambda$ , és szórásnégyzete  $\sigma^2 = \lambda$ . Ezt kihasználva a standardizált normális eloszlás táblázata használható, a 3.3.1.1. fejezetben leírtak szerint.

$$\sigma = \sqrt{37.7} = 6.14$$

$$\text{Próbastatisztika: } \mathbf{u} = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} = \frac{37.7 - 20}{6.14} \cdot \sqrt{3} = \mathbf{4.99}$$

Összehasonlító érték: a standardizált normál eloszlás táblázatából,  $\alpha = 5\%$  egyoldali alternatív hipotézisnél:  $\mathbf{u} = 1.65$ .

Döntés: A számított érték nagyobb, mint a táblázatos érték, ezért nullhipotézisünket elutasítjuk. A felületi mikrobaszám szignifikánsan nagyobb a megengedett értéknél.

### 4.1.3. Normál eloszlású sokaság várható értékének statisztikai próbái

Normál eloszlású sokaság várható értékével kapcsolatos hipotézisvizsgálatoknál attól függően, hogy a sokaság szórása ismert, vagy azt is a mintából kell kiszámítanunk, különböző módon járunk el.

Azokban az esetekben amikor az alapsokaság szórását előzetesen nagy szabadsági fokok mellett meghatároztuk és annak változatlanságáról meg vagyunk győződve, a mintaelemek szórását nem vesszük figyelembe (hacsak nem a szórás ellenőrzése a célunk). Ilyenkor a mintaelemekből csupán az átlagértéket számítjuk ki, és erre vonatkozóan végzünk statisztikai próbát a standardizált normális eloszlás táblázatának felhasználásával.

Amikor az alapsokaság szórása ismeretlen, annak értékét is a mintaelemekből számítva becsüljük. Ebben az esetben a várható értékre vonatkozó hipotézisvizsgálatainkhoz végzett statisztikai próbát a Student féle t táblázat alapján végezzük el.

#### 4.1.3.1. Ismert szórású normál eloszlású sokaság statisztikai próbája. (Egymintás u-próba)

#### 4.4. Példa

Fermentált tejtermék élősejtszáma vonatkozó belső üzemi előírás szerint annak el kell érnie a  $10^8$  sejt/ml értéket. A gyártási tételeken belüli szórás előzetes felmérések alapján ismert, nagysága 0.6 lg egység.

Egy gyártási tétel ellenőrzése során 5 elemű mintát vesznek, melyeknek milliliterenkénti élősejtszámai a következők:  $1.26 \cdot 10^7$ ,  $2.51 \cdot 10^7$ ,  $9.5 \cdot 10^7$ ,  $9.6 \cdot 10^6$ ,  $1.41 \cdot 10^8$ . Kérdés, hogy a mintázott tétel megfelel-e az előírt,  $\mu = 10^8$  sejt/ml várható értéknek.

A mikrobaszámok normális eloszlását logaritmikus transzformációval biztosítjuk, ennek megfelelően a lg N értékek a következők: 7.10, 7.40, 7.98, 6.98, 8.15. Átlagérték: 7.52.

*Nullhipotézis:* A mintázott sokaság várható értéke nem különbözik szignifikánsan az előírt értéktől.

*Alternatív hipotézis:* A tétel várható értéke szignifikánsan kisebb az előírtnál.

*Próbastatisztika:* 
$$u = \frac{\mu - \bar{x}}{\sigma} \cdot \sqrt{n} = \frac{8.00 - 7.52}{0.60} \cdot \sqrt{5} = 1.79$$

*Összehasonlító érték:* a standardizált normál eloszlás táblázatából,  $\alpha = 5\%$  egyoldali alternatív hipotézisnél:  $u = 1.65$ .

*Döntés:* A számított érték nagyobb, mint a táblázatos érték, ezért nullhipotézisünket elutasítjuk. A várható érték szignifikánsan kisebb az előírtnál.

#### 4.1.3.2. Ismeretlen szórású normál eloszlású sokaság statisztikai próbája. (Egymintás t-próba)

#### 4.5. Példa

Az előző pontban ismertetett kérdésfeltevést vizsgáljuk, de a tételen belüli szórást a mintaelemekből számítjuk ki.

A lg N értékek szórása:  $S = 0.522$

*Nullhipotézis:* A mintázott sokaság várható értéke nem különbözik szignifikánsan az előírt értéktől.

*Alternatív hipotézis:* A tétel várható értéke szignifikánsan kisebb az előírtnál.

*Próbastatisztika:* 
$$t = \frac{\mu - \bar{x}}{S} \cdot \sqrt{n} = \frac{8.00 - 7.52}{0.522} \cdot \sqrt{5} = 2.06$$

*Összehasonlító érték:* a Student féle t táblázatából, az  $5-1=4$  szabadsági fokhoz és  $\alpha = 5\%$ -hoz tartozó érték:  $t = 2.13$

*Döntés:* A számított érték kisebb, mint a táblázatos érték, ezért nullhipotézisünket elfogadjuk. A várható érték előírttól való eltérése nem szignifikáns.

### **Kísérlettervezési szempontok**

Összehasonlítva a kétféle döntési eljárást, (4.4. és 4.5. példa) megállapítható, hogy azonos szórásoknál, (ami itt nagyjából teljesül, mert a mintából számított szórás 0.522 közel azonos az állandónak tekintett 0.60 értékkel), a szórást a mintából számítva, kevésbé hatékony a statisztikai próba. Ennek oka, hogy ismert szórás esetén az u táblázatot használjuk, míg mintából számított szórás esetén a t-táblázatot. A t értékek mindig nagyobbak, mint az u értékek (csak 100 fölötti szabadsági fokoknál egyeznek meg közelítőleg). A nagyobb t-értékek következtében az u-próbához képest csak nagyobb különbségek bizonyulnak szignifikánsnak.

A táblázatos t értékek csökkenése a szabadsági fokok növelésével  $Sz_f=10$  felett nagyon lelassul, ezért a párhuzamosok számának növelése jószerivel hatástalan a próba élességére.

## **4.2. KÉTMINTÁS STATISZTIKAI PRÓBÁK**

Kétmintás statisztikai próbáknál mindig két sokaság statisztikai jellemzőit (szórás, várható érték) hasonlítjuk össze. Nullhipotézisünk általában az, hogy az összehasonlítandó jellemzők nem különböznek egymástól szignifikánsan.

### **4.2.1. Két számított szórás összehasonlítása (F-próba)**

Két, mintából számított szórásérték összehasonlítása F-próbával történik. Két szórásnégyzet hányadosa F eloszlást követ, amelyhez a kritikus érték az F táblázatból a számláló és nevező szabadsági fokainak, valamint a választott valószínűségi szintnek megfelelően kikereshető. A hányados képzésénél úgy járunk el, hogy mindig a nagyobb érték kerüljön a számlálóba. A szabadsági fokok számítása:

$Sz_F = n - 1$ , ahol n a mintaelemek száma

### **4.6. Példa**

Lemezöntéses és felületi szélesztéses élősejtszám-meghatározási módszereket összehasonlítva az alábbi eredményeket kapták.

Lemezöntés: 9 párhuzamosból számítva,  $S = 0.102$  lg egység

Szélesztés: 15 párhuzamosból számítva  $S = 0.160$  lg egység



*Nullhipotézis:* a két szórás között nincs szignifikáns különbség

*Alternatív hipotézis:* a két szórás között szignifikáns eltérés van

*Próbastatisztika:*  $F = S_1^2/S_2^2 = 0.160^2/0.102^2 = 2.46$

*Összehasonlító érték:* az F táblázatból, SzF<sub>számláló</sub> = 8, SzF<sub>nevező</sub> = 14, P = 95 %, **F = 2.7**

*Döntés:* A számított érték kisebb, mint a táblázatos érték, ezért a nullhipotézist elfogadjuk. A két módszer szórása közötti szignifikáns különbség nem bizonyítható.

### Kísérlettervezési szempontok

Az F táblázatban összefoglalt értékeket vizsgálva, szembetűnő, hogy az értékek csökkenése sokkal érzékenyebb a nevező szabadsági fokának növelésére. A nevező 6 fölötti szabadsági fokainál ez a csökkenés azonban lelassul. A számláló szabadsági fokának növelése 3 szabadsági fok felett csak igen kis mértékben csökkenti az F értéket. Az F eloszlás ezen tulajdonságának igen nagy szerepe van a célszerű kísérlettervezésben.

Két szórásérték összehasonlításakor a próba hatékonysága annál jobb, minél kisebb az összehasonlító táblázatos érték. Ennek megfelelően a párhuzamosok számát úgy célszerű megválasztani, hogy a számláló és a nevező szabadsági fokainak alapján a legkisebb táblázatos F értéket kapjuk. A szabadsági fokok helyes megválasztásának különösen kis szabadsági fokok esetén van fontos szerepe. Nagyobb szabadsági fokoknál ez a hatás csökken.

## 4.2.2. Poisson eloszlású sokaságok várható értékének összehasonlítása

### 4.7. Példa

Két táptalajt kívánunk összehasonlítani kis mikrobaszámok kimutatása tekintetében. Az azonos mintából történő meghatározások során a Petri csészéken kifejlődött telepek száma:

1. táptalaj:	5, 8, 3, 7, 2	Összesen	$k_1 = 25$	$n_1 = 5$
2. táptalaj:	9, 6, 8, 11, 3, 7,	Összesen	$k_2 = 44$	$n_2 = 6$

A Petri csészéken kifejlődött telepek száma kicsiny, Poisson eloszlást követ.

*Nullhipotézis:* a két táptalajon kinőtt telepek számának várható értéke között nincs szignifikáns eltérés

*Alternatív hipotézis:* a várható értékek között szignifikáns különbség van

*Próbastatisztika:*  $\chi^2 = \left( \frac{k_1}{n_1} - \frac{k_2}{n_2} \right)^2 \cdot \frac{n_1 \cdot n_2}{k_1 + k_2} = 2.36$

*Összehasonlító érték:*  $\chi^2$  táblázatból  $\alpha = 5\%$ , Szf. = 1  $\chi^2 = 3.84$

*Döntés:* A számított érték kisebb a táblázatos értéknél, ezért nullhipotézisünket elfogadjuk. A két táptalajon meghatározott telepszámok várható értéke között nincs szignifikáns különbség.

### 4.2.3. Normál eloszlású sokaságok várható értékének összehasonlítása

#### 4.2.3.1. Ismert szórású sokaságok várható értékének összehasonlítása (Kétmintás u-próba)

##### 4.8. Példa

Fermentált tejtermék két gyártási tételének élősejtszámát hasonlítjuk össze. A milliliterenkénti élősejtszám tétel belüli szórása előzetes felmérések alapján ismert:  $\sigma = 0.60$  lg egység. Az összehasonlító vizsgálatokat a normális eloszlás biztosítása érdekében a sejtszámok logaritmusával végezzük. A kapott eredmények:

1. tétel:	$n_1 = 5$	Átlagérték:	$\lg N_1 = 8.30$
2. tétel:	$n_2 = 5$	Átlagérték:	$\lg N_2 = 7.85$

*Nullhipotézis:* a két gyártási tétel élősejtszámának várható értéke között nincs szignifikáns eltérés

*Alternatív hipotézis:* a várható értékek között szignifikáns különbség van

*Próbastatisztika:* 
$$u = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Esetünkben a szórások megegyeznek:  $\sigma_1 = \sigma_2 = \sigma$ , ezért a próbastatisztika:

$$u = \frac{\bar{x}_1 - \bar{x}_2}{\sigma} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} = 1.19$$

*Összehasonlító érték:* u táblázatból  $\alpha = 5\%$ , egyoldali **u = 1.65**

*Döntés:* A számított érték kisebb a táblázatos értéknél, ezért nullhipotézisünket elfogadjuk. A két gyártási tétel élősejtszámának várható értéke között nincs szignifikáns különbség.

#### 4.2.3.2. Ismeretlen szórású sokaságok várható értékének összehasonlítása

Ismeretlen szórású sokaságok várható értékének összehasonlításakor a szórások értékét a mintákból számítjuk ki. Amennyiben a két minta szórása között nincs szignifikáns különbség (F-próba), a várható értékek összehasonlításának statisztikai próbáját a két szórásból számított közös szórással végezzük el. A kiértékeléshez használt t eloszlás táblázatos értékének kiválasztásakor ilyenkor a szabadsági fok a két szabadsági fok összege.

Ha az F próba szignifikáns különbséget mutat ki a két szórás között, akkor a számítás a t érték szabadsági fokának meghatározása miatt kissé bonyolultabb.

#### Azonos szórású sokaságok összehasonlítása (Kétmintás t-próba)

#### 4.9. Példa

Nyerstej tételek elősejtszám meghatározását két módszer felhasználásával végezzük el. Az azonos mintákból meghatározott sejtszámok logaritmusait a 4./1. Táblázatban foglaltuk össze. Arra vagyunk kíváncsiak, hogy a két módszerrel meghatározott sejtszámok várható értékei között van-e szignifikáns különbség.

#### 4./1. Táblázat. Nyerstej tételek mikrobiológiai vizsgálatának eredményei (lg N/ml)

Minta	A	B.	d=B-A
1.	5.47	5.60	0.13
2.	5.30	5.40	0.10
3.	6.00	6.04	0.04
4.	5.47	5.49	0.02
5.	5.14	5.20	0.06
6.	4.98	5.04	0.06
7.	5.30	5.38	0.08
8.	5.04	5.14	0.10
9.	5.47	5.52	0.05
10.	5.16	5.20	0.04
<b>Átlag</b>	5.333	5.401	0.068
<b>S</b>	0.2946	0.2886	0.0339
<b>n</b>	10	10	10

A két módszer szórását F próbával hasonlítjuk össze:  $F = 0.2946^2 / 0.2886^2 = 1.04$

Táblázatos érték:  $\alpha = 5\%$       Szf.<sub>számláló</sub> = 9      Szf.<sub>nevező</sub> = 9      F = 3.1

A számított érték kisebb a táblázatosnál ezért a szórások között nincs szignifikáns különbség. A továbbiakban közös szórás feltételezésével számolhatunk.

*Nullhipotézis:* A két módszer által adott eredmények várható értéke között nincs szignifikáns különbség

*Alternatív hipotézis:* A két várható érték szignifikánsan különbözik.

*Próbastatisztika:*  $t = \frac{|\bar{x}_1 - \bar{x}_2|}{S_d}$ ,

ahol 
$$S_d = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.1375$$

*Próbastatisztika számított értéke:*  $t = 0.495$

**Összehasonlító érték:** t táblázatból Szf. = 18,  $\alpha = 5\%$  kétoldali **t = 2.10**

**Döntés:** A számított érték kisebb, mint a táblázatos érték, ezért a nullhipotézist elfogadjuk. A két módszerrel meghatározott sejtszámok várható értéke között nincs szignifikáns különbség.

A kiértékelést matematikai-statisztikai programcsomaggal (STATGRAPHIC 5.1) elvégezve a várható értékek közötti különbség szignifikancia szintje: 0.608, ami azt jelenti hogy csak 60.8%, vagy ennél nagyobb elsőfajú hiba esetén bizonyul az eltérés szignifikánsnak.

### ***Különböző szórású sokaságok összehasonlítása (Welch-próba)***

Ha az alapsokaságok szórása ismeretlen és a két mintából meghatározott  $S_1$  és  $S_2$  szórás egymástól szignifikánsan különbözik, a t próba nem alkalmazható. A várható értékek statisztikai próbája közelítőleg Welch módszerével oldható meg.

**Nullhipotézis:** A két sokaság várható értéke nem különbözik szignifikánsan

**Alternatív hipotézis:** A várható értékek közötti eltérés szignifikáns.

**Próbastatisztika:** 
$$t_f = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

A táblázatos t érték f szabadsági fokának meghatározása az alábbi összefüggések alapján történik:

$$\frac{1}{f} = \frac{1}{n_1 - 1} \left[ \frac{\frac{S_1^2}{n_1}}{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right]^2 + \frac{1}{n_2 - 1} \left[ \frac{\frac{S_2^2}{n_2}}{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right]^2$$

**Összehasonlító érték:** t táblázatból az f szabadsági fokhoz és a választott valószínűségi szinthez tartozó érték.

### ***4.2.3.3. Párosított adatok eltérésére vonatkozó statisztikai próba. (Páros t-próba)***

#### ***4.10. Példa***

Nyerstej mikrobaszámának meghatározására szolgáló két módszert kívánunk összehasonlítani. A különböző minták két módszerrel meghatározott élősejtszámainak logaritmusait, valamint a két módszer közötti különbségeket a 4./1. Táblázat tartalmazza. A matematikai statisztikai kiértékelést a d értékek egymintás t-próbájával végezzük.

A d értékek statisztikai jellemzői: Átlag = 0.068, Szórás = 0.0339, n = 10

*Nullhipotézis:* A két módszer által meghatározott mikrobaszámok különbségének várható értéke nem tér el szignifikánsan 0-tól.

*Alternatív hipotézis:* A különbségek várható értéke szignifikánsan eltér 0-tól.

*Próbastatisztika:*  $t = \frac{\bar{d}}{S} \sqrt{n} = 6.34$

*Összehasonlító érték:* t táblázatból  $\alpha = 5\%$ , kétoldali Szf = 9 **t = 2.26**

*Döntés:* A számított érték nagyobb, mint a táblázatos érték, ezért a nullhipotézist elutasítjuk. A két módszer által meghatározott sejtszámok között szignifikáns eltérés van.

A B-módszerrel meghatározott lgN értékek átlagosan 0.068 lg egységgel nagyobbak. Ezt átszámolva normál sejtszámmra:  $10^{0.068} = 1.17$ -szer, azaz 17%-kal nagyobb sejtszám határozható meg a B-módszerrel, mint az A-módszerrel.

### **Kísérlettervezési szempontok**

Az előző, 4.9. példában a két módszer összehasonlításakor a vizsgált sokaságok várható értékei között nem tudtunk szignifikáns különbséget kimutatni. Ennek oka az, hogy a vizsgált sokaságok eloszlása teljességgel átfedte egymást, s ezáltal eltakarta a két módszer között lévő szisztematikus eltérést. A vizsgált sokaságok eloszlásának zavaró hatása kiküszöbölhető, ha nem a teljes eloszlásokat hasonlítjuk egymáshoz, hanem az azonos mintákból végzett meghatározások különbségének eloszlását vizsgáljuk.

Képezve az összetartozó mérési eredmények különbségét:  $d_i = x_{Bi} - x_{Ai}$ , ezeknek a különbségeknek az átlagértéke normális eloszlást követ, melynek várható értéke, - ha a módszerek között nincs szignifikáns eltérés, - 0, szórása a  $d_i$  értékekből számítható

### **4.3. TÖBBMINTÁS STATISZTIKAI PRÓBÁK**

Többmintás statisztikai próbákkal kettőnél több sokaság statisztikai jellemzőit (gyakoriság, szórás várható érték) hasonlítjuk össze. Nullhipotézisünk, hasonlóan az egy- és kétmintás próbákhoz, általában az, hogy a vizsgált sokaságok összehasonlítandó paraméterei között nincs szignifikáns eltérés. Amennyiben a próba eredménye ennek ellentmond, az azt jelenti, hogy legalább az egyik sokaság vizsgált paramétere szignifikánsan eltér a többitől. A többmintás statisztikai próbák, – különösen a várható értékek összehasonlítását lehetővé tevő varianciaanalízisek esetében, – a különbségek kimutatása mellett több tényező együttes hatásának vizsgálatát is lehetővé teszik, ezért a kísérlettervezés és –kiértékelés leghatékonyabb eszközeinek bizonyultak.

### 4.3.1. Poisson eloszlású gyakoriságok összehasonlítása

#### 4.11. Példa

Üdítőital tételek élesztőgombás szennyezettségét vizsgálva membránszűréses eljárással, a 4./2. Táblázatban összefoglalt eredményeket kapták. Kérdés, hogy a gyártás folyamán az átlagos élesztőgombaszám változása szignifikáns-e.

#### 4./2. Táblázat Üdítőital élesztőgombás szennyezettségének változása (sejt/100 ml)

	Tételek					
	1.	2.	3.	4.	5.	6.
	8	2	6	15	3	4
	4	6	9	16	2	6
	10	3	7	20	6	4
$n_i$	3	3	3	3	3	3
$k_i$	22	11	22	51	11	14
$\lambda_i = k_i/n_i$	7.33	3.67	7.33	17.00	3.67	4.67
$(\lambda_i - \lambda)^2$	0.0025	13,031	0,0025	94,4784	13,0321	6,8121
$n_i \cdot (\lambda_i - \lambda)^2$	0.0075	39.0963	0.0075	283.4352	39.0963	20.4363

Jelölések:  $n_i$  = mintánkénti párhuzamosok száma = 3

$k_i$  = mintánkénti összes sejtszám

$m$  = tételek száma = 6

$\Sigma n_i \cdot (\lambda_i - \lambda)^2 = 382.0791$

$$K = \sum_{i=1}^{i=m} k_i = 131$$

$$N = \sum_{i=1}^{i=m} n_i = 18$$

$$\lambda = K/N = 7.28$$

*Nullhipotézis:* Az üdítőitalok élesztőgombás szennyezettsége között nincs szignifikáns különbség

*Alternatív hipotézis:* Legalább az egyik tétel szignifikánsan eltér a többitől

*Próbastatisztika:* 
$$\chi^2 = \frac{\sum_{i=1}^{i=m} n_i \cdot \left( \frac{k_i}{n_i} - \frac{K}{N} \right)^2}{\frac{K}{N}} = 382.0791/7.28 = \mathbf{52.48}$$

*Összehasonlító érték:*  $\chi^2$  táblázatból  $\alpha = 5\%$  Szf = m-1=5  $\chi^2 = \mathbf{11.1}$

*Döntés:* A számított érték nagyobb a táblázatos értéknél, ezért a nullhipotézist elutasítjuk. Az üdítőital tételek között legalább egy (de lehet, hogy több is) szignifikánsan eltér a többitől.

### 4.3.2. Szórások összehasonlítása

Az összehasonlítandó sokaságok szórásának homogenitása nagyon sokszor alapfeltétele az elvégzendő statisztikai próbáknak, ezért több sokaság összehasonlításánál ezekre a vizsgálatokra mindig sor kerül.

#### *Szórások homogenitásvizsgálata Bartlett-próbával*

Általánosan alkalmazható eljárás normális eloszlású sokaságok szórásának vizsgálatára. Számításmenete kissé bonyolult, de a matematikai-statisztikai programcsomagok elterjedtsége miatt ennek már nincs jelentősége.

#### 4.12. Példa

Pasztörtej tételek mikrobaszámának szórás-vizsgálata során a 4.3. Táblázatban összefoglalt eredményeket kaptuk.

#### **4.3. Táblázat.** Pasztörtej mikrobaszámának tételen belüli szórásai (lg N/ml)

Tétel	1.	2.	3.	4.	5.	$\Sigma$
$S_i$	0.204	0.186	0.160	0.195	0.210	
$n_i$	3	5	4	3	5	20
$S_i^2$	0.0416	0.0346	0.0256	0.0380	0.0441	
$(n_i-1)S_i^2$	0.0382	0.1384	0.0768	0.0761	0.1764	0.5509
$\lg S_i^2$	-1.3809	-1.4609	-1.5918	-1.4202	-1.3556	
$(n_i-1)\lg S_i^2$	-2.7618	-5.8436	-4.7754	-2.8404	-5.4224	-21.6472

$$S^2 = \frac{\sum_{i=1}^m (n_i - 1) S_i^2}{\sum_{i=1}^m (n_i - 1)} = 0.5509/15 = 0.0367 \quad \lg S^2 = -1.4353$$

$$C = 1 + \frac{1}{3 \sum_{i=1}^m (n_i - 1)} \left[ \sum_{i=1}^m \frac{1}{n_i - 1} - \frac{1}{\sum_{i=1}^m (n_i - 1)} \right] = 1 + \frac{1}{45} [1.8333 - 0.0667] = 1.0393$$

$$\chi^2 = \frac{2.30}{C} \left[ \sum_{i=1}^m (n_i - 1) \cdot \lg S^2 - \sum_{i=1}^m (n_i - 1) \lg S_i^2 \right] = \frac{2.30}{1.0393} [-15 \cdot 1.4353 + 21.6472] = 0.26$$

C értéke 1-nél nagyobb, ezért csak abban az esetben érdemes kiszámolni, ha a  $\chi^2$  érték C=1 behelyettesítéssel nagyobbak adódik a táblázatos értéknél. C pontos kiszámításával  $\chi^2$  esetleg a kritikus érték alá csökkenthető.

*Nullhipotézis:* A szórások homogének, nincs köztük szignifikáns különbség.

*Alternatív hipotézis:* A szórások közül legalább egy szignifikánsan különbözik

*Próbastatisztika:*  $\chi^2 = 0.26$

Összehasonlító érték:  $\chi^2$  táblázatból  $\alpha = 5\%$ , Szf = m-1 = 4  $\chi^2 = 9.49$

*Döntés:* A számított érték kisebb a táblázatosnál, ezért a nullhipotézist elfogadjuk. Az összehasonlított tételek szórása homogénnek tekinthető. Közös értékük az  $S^2$  négyzetgyökeként adható meg:  $S = 0.1916$  lg egység.

### 4.3.3. Várható értékek összehasonlítása variancia-analízissel

Kettőnél több minta várható értékének összehasonlítására szolgáló matematikai statisztikai eljárások, melyek a leghatékonyabb kiértékelést teszik lehetővé. A variancia analízis számításmenete a következő lépésekből áll.

1. Az eredményekben mutatkozó eltéréseket két részre osztjuk

- Véletlenre visszavezethető eltérés (kísérleti hiba)
- Kezelések hatására visszavezethető eltérés.

A különböző eredetű eltéréseket szórásnégyzet (variancia) formájában adjuk meg, innen ered az eljárás neve.

2. A kezelések hatására visszavezethető eltéréseket (szórásnégyzetek) viszonyítjuk a véletlen okozta szórásnégyzethez. Ennek az összehasonlításnak alapfeltételei a következők:

- Az eredmények a kezelésektől eltekintve függetlenek legyenek egymástól.
- Az egyes kezeléseken belül az eredmények eloszlása normális legyen.
- A kezeléseken belüli véletlen szórásnégyzetek között ne legyen szignifikáns különbség.

3. Ha a kezelések hatására visszavezethető szórásnégyzet szignifikánsan nagyobb a véletlen szórásnégyzethöz, a kezelések hatása további csoportosítások alapján elemezhető.

Ha az alapadatokra vonatkozóan nem teljesül a 2. pontban összefoglalt feltételek bármelyike, az adatok csak megfelelő transzformáció (pl. sejt számok logaritmikus transzformációja) után használhatók fel a számításokhoz.

- Az adatok függetlensége megfelelő kísérleti elrendezéssel biztosítható
- A normális eloszlás teljesülését a párhuzamosok növelése a centrális határeloszlás következtében többnyire biztosítja. 3 párhuzamostól kezdve az átlagértékek eloszlása már igen nagy valószínűséggel normálisnak tekinthető.
- Legnagyobb problémát a kezeléseken belüli szórások eltérése okozhat, ezért a szórások homogenitásáról minden esetben előzetesen meg kell győződni.

A variancia-analízis segítségével megoldható feladatokat két nagy csoportra oszthatjuk:

- Több azonos szórású, normális eloszlású valószínűségi változó várható értékének összehasonlítása. (Egyszempontos variancia-analízis).
- A vizsgált változó értékét szignifikánsan befolyásoló hatások kiválasztása. (Egy- és többszempontos variancia-analízis).



Ha egy hatás (faktor), amely a várható értéket szignifikánsan befolyásolja, önmagában is mennyiségi változó (pl. hőmérséklet, vegyszerkoncentráció, kezelési idő, stb.), akkor a variancia-analízist követően összefüggés-vizsgálatokra is van lehetőség. Ezekben az összefüggésekben a szignifikáns hatású faktor különböző értékei képviselik a független változót. Előfordulnak olyan feladatok, amelyekben az egyes faktorok nem mennyiségi (pl. különböző kezelési eljárások), mások azonban mennyiségi jellemzők (pl. fertőtlenítőszer koncentráció). A kétféle hatás együttes értékelését kovariancia-analízissel végezhetjük el.

Számítástechnikailag a variancia-analízis során az összes adatból számított négyzetösszeget felbontjuk az egyes faktorok szerint csoportosított négyzetösszegekre. Faktorok szerint kiszámítva az egyes négyzetösszegeket, az összes adatból számított és a faktorok szerinti négyzetösszegek különbsége adja a kölcsönhatásokra utaló tagot, amely egyes esetekben a véletlen hibát reprezentálja. A négyzetösszegekből a hozzájuk tartozó szabadsági fokokkal való osztás révén szórásnégyzeteket számolunk és ezeket viszonyítjuk a véletlen szórásnégyzethez. A hatás szignifikanciáját F-próbával döntjük el.

Ha az a cél, hogy a vizsgált folyamatban a tiszta hatások mellett a kölcsönhatásokat is felderítsük, akkor a kölcsönhatásokból számított szórásnégyzetet mindig a véletlen, vagyis a párhuzamosok között lévő szórásnégyzethez viszonyítjuk.

A kölcsönhatások értelmezése minden esetben szakmai probléma. Amennyiben a kölcsönhatás a folyamat velejárója (pl. a különböző helyről származó nyersanyagok mikroflórája eltérő módon reagál az egyes technológiai lépésekre), célszerű a véletlen hibával összevonni a kölcsönhatások következtében fellépő eltéréseket is, illetve ezt a kölcsönhatási tagot tekinteni a véletlen hibának.

Ha a véletlen okozta szórásnégyzetet nem párhuzamos vizsgálatok eredményeiből számítjuk, hanem a párhuzamosok átlagértékeivel végzett variancia-analízis maradék szórásnégyzetével becsüljük, akkor az egyedi értékekre vonatkozó szórásnégyzet az átlagértékek szórásnégyzetének a párhuzamosok számával való szorzással számítható.

### **Kísérlettervezési szempontok.**

A véletlen hiba meghatározásának legmegbízhatóbb módja, hogy a vizsgálatokat mindig legalább 2 párhuzamosban végezzük el. Ebben az esetben a véletlen szórásnégyzet szabadsági foka  $k \cdot (p-1)$ , ahol  $k$  a kezelések,  $p$  pedig a kezeléson belüli párhuzamosok száma. Mivel az F-próba érzékenysége a nevező 10 feletti szabadsági fokainál lecsökken, 5-nél több kezelés esetén a kezeléseken belüli párhuzamosok számát nem érdemes 3-nál nagyobbra választani.

Ha nincsenek párhuzamosok, a véletlen hatást a független hatások levonása után maradó szórásnégyzettel becsüljük. Ez a módszer az előzőhöz képest minden esetben információvesztést okoz, mert nem teszi lehetővé a kölcsönhatások kimutatását.

Párhuzamosok tekintetében leghatékonyabb a kísérleti elrendezés, ha azok száma kezeléseken belül megegyezik. Egyszempontos variancia-analízis eltérő párhuzamosokkal is kiértékelhető, többszempontos variancia-analízisnél azonban feltétel a kezeléseken belüli azonos párhuzamos szám.

A variancia-analízis mintapéldák megoldásánál csak az alapösszefüggéseket ismertetjük, a részletes számításokat STATGRAPHIC 5.1 statisztikai programmal végezzük el. Tekintettel arra, hogy napjainkban a hasonló jellegű feladatok megoldása mindenütt számítógéppel történik, a továbbiakban nem foglalkozunk a kézzel végzett számítások megkönnyítését célzó adategyszerűsítéssel. Csupán egy nagyon fontos szempontra kívánjuk felhívni a figyelmet: Az alapadatok ésszerű kerekítése megengedett, de a részeredmények kerekítésénél mindenkor figyelembe kell venni a számításmenet által megkövetelt pontosságot. Sok esetben nagy számok igen kis különbségével számolunk, ezért a tizedesek elhagyása hibát okozhat.

### 4.3.3.1. Egyszempontos variancia-analízis

#### 4.13. Példa

Különböző gyártási tételből származó biojoghurtok bifidobaktérium-számát hasonlítjuk össze. Kérdés, hogy van-e különbség az egyes tételek várható értéke között. A mérési eredményekből számított lg sejszámokat a 4./4. Táblázat tartalmazza.

#### 4./4. Táblázat. Különböző gyártásból származó biojoghurtok bifidobaktérium-számai (lg N/ml)

	Tétel				
	A	B	C	D	E
	8.12	7.98	8.00	6.50	7.68
	8.04	7.63	7.96	6.30	7.40
	8.30	7.76	7.30	6.00	7.30
	8.00	7.54	7.64	6.26	7.80
Átlag	8.115	7.728	7.728	6.265	7.545

A kezelésen belüli szórások homogének, ezért a kiértékelés variancia-analízissel elvégezhető.

n = összes adatszám            n = 20  
k = kezelések száma            k = 5  
p = párhuzamosok száma        p = 4

A variancia-analízishez az alábbi négyzetösszegek kiszámítása szükséges, melyek alapján a Variancia-táblázat kitölthető.

$$a = \sum_{i=1}^n x_i^2, \quad b = \frac{\left[ \sum_{i=1}^n x_i \right]^2}{n}, \quad c = \sum_{i=1}^k \frac{\left[ \sum_{j=1}^p x_{ij} \right]^2}{p}$$

A varianci-táblázat variancia-forrás oszlopában zárójelben feltüntettük a négyzetösszeg és szabadsági fok számítási módját. Az  $S^2$  és az F érték kiszámítása:

$$S^2 = \text{Négyzetösszeg} / \text{Szabadsági fok}, \quad F = S^2 / S_0^2$$

A véletlen szórásnégyzet a kezelésen belüli párhuzamosok szórására jellemző közös szórásnégyzet, ami tulajdonképpen az egyes kezelésekre jellemző szórásnégyzetek átlaga.

A négyzetösszegek és a szabadsági fokok additívek (összeadhatók, kivonhatók). A maradék (véletlen) tag négyzetösszege és szabadsági foka ezért úgy is számítható, hogy az összesből kivonom a kezelésekre számított megfelelő értéket. A maradék négyzetösszeget elosztva a hozzátartozó szabadsági fokkal, eredményül kapjuk a véletlen szórásnégyzetet ( $S_0^2$ ).

A véletlen szabadsági fok kiszámításának másik módja:  $Szf = k \cdot (p-1)$ . A kétféle módon számított értéknek természetesen egyeznie kell.

**4./5. Táblázat.**

## Variancia-táblázat

Variancia forrás	Négyzetössze g	Sz.. fok	S <sup>2</sup>	F	Szign. szint
Összes (a-b)	8.7917	20-1=19			
Kezelések között (c-b)	8.0194	5-1=4	2.00	38.9***	0.0000
Véletlen (a-c)	0.7723	15	0.0515 = S <sub>0</sub> <sup>2</sup>		

Az F-próba kritikus értékei a számláló 4, a nevező 15 szabadsági fokánál:  $F_{95\%} = 3.1$ ,  $F_{99\%} = 4.9$

A számított érték ennél jelentősen nagyobb, ezért azt mondjuk, hogy az összehasonlított sokaságok várható értékei között szignifikás különbség van.

A szignifikancia szint jelölésére a számított F érték után tett csillagok utalnak. Egy csillag 95, két csillag 99, három csillag 99.9%-os valószínűségű szignifikás eltérésre utal.

Esetünkben a számítógépes program által megadott szignifikancia-szint 0.0000, ami azt jelenti, hogy a szignifikás különbség valószínűsége 99.99% feletti (0.01% elsőfajú hiba megengedése már szignifikás eltérést eredményez).

Az átlagértékek közötti legkisebb szignifikás különbség (LSD) számítása:

$$LSD = t \cdot \sqrt{S_0^2 \cdot \frac{2}{p}} = t \cdot S_0 \cdot \sqrt{\frac{2}{p}},$$

ahol  $S_0^2$  a véletlen szórásnégyzet, p pedig az összehasonlítandó kezeléseken belüli párhuzamosok száma. A táblázatos t érték szabadsági foka mindig a véletlen szórásnégyzet szabadsági foka.

A 15-ös szabadsági fokhoz és 95% kétoldali valószínűségi szinthez tartozó érték:  $t = 2.13$ .

$$S_0 = \sqrt{0.0515} = 0.227$$

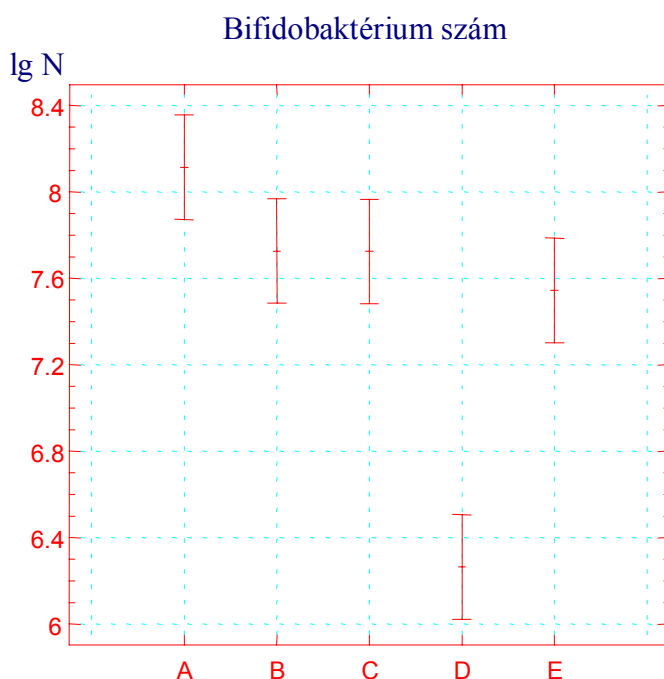
Az átlagértékek közötti legkisebb szignifikás differencia:  $LSD = 2.13 \cdot 0.227 \cdot \sqrt{2/4} = \mathbf{0.3419}$

Sorrendbe állítva a kezelésösszegeket és különbségeiket összehasonlítva az LSD értékkel:

Kezelés	Átlagérték	Homogenitás
D	6.265	*
E	7.545	*
C	7.728	*
B	7.728	*
A	8.115	*

Az egy oszlopban lévő csillagok homogenitásra utalnak, míg az eltérő oszlopban lévők a várható értékek közötti szignifikáns eltérést jelzik.

A variancia-analízis eredményeit a 4./1. ábra szemlélteti, amelyen a különböző kezelésekhez tartozó átlagértékeket és azok közös konfidencia-intervallumát tüntettük fel. A konfidencia-intervallumok esetenként átfedhetnek egymást, de ez még nem jelenti azt, hogy nincs szignifikáns különbség a két átlagérték között. A legkisebb szignifikáns különbség számított értéke az átlagértékek közötti különbségra vonatkozik és nem a konfidencia-határookra.



**4./1. ábra**

Biojoghurtok bifidobaktérium-szám várható értékeinek konfidencia-intervalluma.

#### 4.3.3.2. Kétszemponos variancia-analízis

Két lehetséges kísérleti elrendezéssel foglalkozunk. A kevesebb mintaszámot igénylő ismétlés nélküli esetben az egyes kezeléseken belül nem alkalmazunk párhuzamosokat. Az ismétléses esetekben a kezelésen belüli párhuzamosok miatt a mintaszám többszörösére nő, viszont lehetőségünk van a kölcsönhatások felderítésére, illetve a véletlen hiba kölcsönhatástól mentes meghatározására.

##### 4.3.3.2.1. Kétszemponos variancia-analízis ismétlések nélküli elrendezéssel

#### 4.14. Példa

A gyártástechnológia kritikus pontjainak feltárása érdekében a mezofil aerob spóraszám alakulását vizsgálták paradicsom-sűrítmenyt előállító vonalon. A különböző gyártási tételek esetében az egyes technológiai pontokon meghatározott értékeket a 4./6. Táblázatban foglaltuk össze. A kiértékelés során a gyártás kritikus pontjainak felderítése volt a cél, vagyis meghatározni azokat a technológiai lépéseket, ahol az aerob spóraszám szignifikánsan megváltozik. E mellett vizsgálni kívánták az egyes gyártási tételek közötti eltérések szignifikanciáját is.

**4./6. Táblázat.** Paradicsom-sűrítmeny gyártási fázisainak aerob spóraszámai (lg N/ml)

Gyártás	Fázisok				Összeg
	A	B	C	D	
1.	3.89	2.86	2.36	1.78	10.89
2.	3.67	3.36	4.00	2.00	13.03
3.	3.79	3.71	3.49	2.49	13.48
4.	3.42	4.40	3.00	2.38	13.20
5.	3.60	4.24	4.38	1.38	13.60
<b>Összeg.</b>	18.37	18.57	17.23	10.03	64.20
<b>Átlag</b>	3.674	3.714	3.446	2.006	3.210

n = összes adatszám            n = 20  
k = gyártási tételek száma    k = 5  
f = fázisok száma                f = 4

$$a = \sum_{i=1}^n x_i^2 = 221.0638 \qquad b = \frac{\left[ \sum_{i=1}^n x_i \right]^2}{n} = 206.0820$$

A variancia-analízisben minden csoportosítás négyzetösszegének a és b közé kell esnie. Amennyiben nem így van, számítási hiba, vagy kerekítési hiba történt.

Fázisok hatása: 
$$c = \sum_{i=1}^4 \frac{\left[ \sum_{j=1}^5 x_{ij} \right]^2}{5} = (18.37^2 + 18.57^2 + 17.23^2 + 10.03^2) / 5 = 215.9552$$

Tételek hatása: 
$$d = \sum_{i=1}^5 \frac{\left[ \sum_{j=1}^4 x_{ij} \right]^2}{4} = (10.86^2 + 13.03^2 + 13.48^2 + 13.20^2 + 13.60^2) / 4 = 207.3208$$

**4./7. Táblázat.** Variancia-táblázat.

Variancia forrás	Négyzetössze g	Sz.. fok	S <sup>2</sup>	F	F <sub>99%</sub>
1. Összes (a-b)	14.9818	20-1=19			
2. Fázisok hatása (c-b)	9.8732	4-1=3	3.29	10.2**	6.0

3. Tételek hatása (d – b)	1.2388	5-1=4	0.3097	<1	
Maradék (1.-2.-3)	3.8698	12	0.3225 = S <sub>0</sub> <sup>2</sup>		

A Varianciátáblázatban a fázisok hatását értékelő F érték szignifikáns különbségre utal. A tételek hatását reprezentáló F értéke 1-nél kisebb, tehát nincs eltérés a gyártási tételek hatása között.

A fázisátlagok legkisebb szignifikáns eltéréseinek számítása, tekintettel arra, hogy a párhuzamosok száma p = 1:

$$S_0 = 0.568 \quad \text{LSD} = t \cdot S_0 \cdot \sqrt{2}$$

A táblázatos t szabadsági foka 12,  $\alpha = 5\%$  egyoldali, mert a D kezelésre gyanakszunk:  $t = 1.78$

$$\text{LSD} = 1.78 \cdot 0.568 \cdot \sqrt{2} = 1.430.$$

Sorrendbe állítva a kezelésátlagokat:

Kezelés	Átlagérték	Homogenitás
D	2.006	*
C	3.446	*
A	3.674	*
B	3.714	*

A gyártás folyamán a spóraszám egészen az utolsó fázisig változatlan marad. A D fázis azonban szignifikánsan csökkenti a spóraszámot.

#### 4.3.3.2.2. Kétszemponos variancia-analízis kezelésen belüli párhuzamosokkal

##### 4.15. Példa

Nyerstej kóliformszámának kimutatására szolgáló táptalajok összehasonlító vizsgálatát végezték el különböző nyerstej-mintákon. A vizsgálatok célja az összehasonlítandó táptalajok közötti esetleges különbségek, valamint a tejminták mikroflórájával való kölcsönhatás kimutatása volt. A meghatározott kóliformszámok logaritmusait a kezelésekre függvényében a 4./8. Táblázatban foglaltuk össze.

##### Kísérlettervezési szempontok

A kezelésen belüli párhuzamosok alkalmazásának célja a véletlen hiba korrekt meghatározása és a táptalajok, valamint a minták közötti esetleges kölcsönhatások kimutatása. Gyakorlatilag előfordulhat, hogy a különböző eredetű minták mikroflórája eltérő összetételű, s egyes táptalajok a coliform mikrobák különböző csoportjaira nézve eltérő szelektivitásuk lehetnek. Hasonló eredményre vezethet a mintákban esetleg jelenlévő gátlószerek hatása is. Ezért táptalaj-összehasonlító kísérletekben mindig kezelésen belüli párhuzamosokkal célszerű dolgozni. A párhuzamosok számára vonatkozóan érvényes az F-próba kritikus értékének 10

szabadsági fok feletti lassuló csökkenése, ezért 10 feletti kezelés esetén nem érdemes 3-nál több párhuzamossal dolgozni. (Tulajdonképpen 2 is elég lenne).

**4./8. Táblázat.**

Nyerstej minták kóliformszáma (lg N/ml).

Tejminta	Táptalaj		
	A	B	C
1.	4.30	4.28	4.42
	4.16	4.16	4.52
	4.20	4.30	4.36
2.	2.18	2.20	2.36
	2.30	2.04	2.40
	2.32	2.24	2.48
3.	3.16	3.24	3.52
	3.21	3.08	3.46
	3.30	3.18	3.30
4.	3.28	3.30	3.56
	3.16	3.32	3.62
	3.12	3.20	3.50

n = összes adatszám            n = 36  
 k = kezelések száma            k = 12  
 l = táptalajok száma            l = 3  
 m = tejminták száma            m = 4  
 p = párhuzamosok száma        p = 3

$$a = \sum_{i=1}^n x_i^2$$

$$b = \frac{\left[ \sum_{i=1}^n x_i \right]^2}{n}$$

$$c = \sum_{i=1}^k \frac{\left[ \sum_{j=1}^p x_{ij} \right]^2}{p}$$

$$d = \sum_{i=1}^l \frac{\left[ \sum_{j=1}^{m \cdot p} x_{ij} \right]^2}{m \cdot p}$$

$$e = \sum_{i=1}^m \frac{\left[ \sum_{j=1}^{l \cdot p} x_{ij} \right]^2}{l \cdot p}$$

**4./9. Táblázat.**

Variancia-táblázat.

Variancia forrás	Négyzetösszeg	Sz.. fok	S <sup>2</sup>	F	Szign. szint
1. Összes (a-b)	19.0588	n-1=35			
2. Véletlen (c-b)	0.1546	12·2=24	0.00644 = S <sub>0</sub> <sup>2</sup>		
3. Táptalajok hatása (d-b)	0.4633	3-1=2	0.2317	36***	0.0000
4. Tejminták hatása (e-b)	18.3848	4-1=3	6.128	951***	0.0000

Kölcsönhatás (1-2-3-4)	0.0561	2·3=6	0.0093	1.45	0.2376
------------------------	--------	-------	--------	------	--------

Az F-próba alapján mindkét vizsgált hatás szignifikáns, viszont nincs kölcsönhatás a táptalajok és a tejminták között.

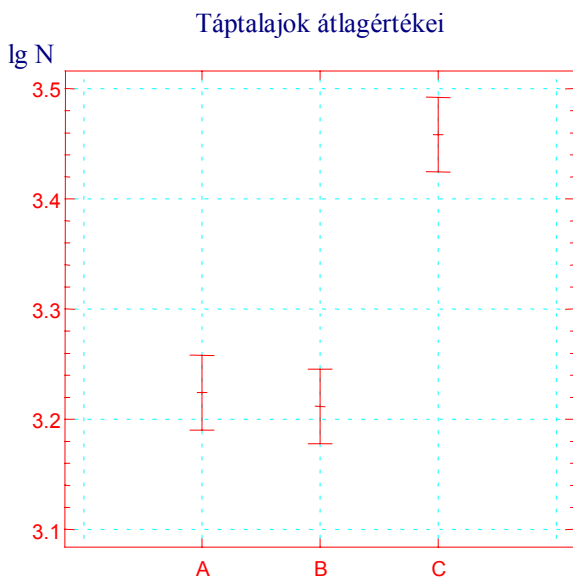
A módszerekre jellemző közös szórásérték:  $S_0 = \sqrt{0.00644} = 0.080$  lg egység

A kezelésközvetlen közötti legkisebb szignifikáns eltérések számításához közös t értéket használhatunk, melynek szabadsági foka a véletlen szóráséval megegyezően 24. Kétoldali 5%-os elsőfajú hibát választva,  $t = 2.06$

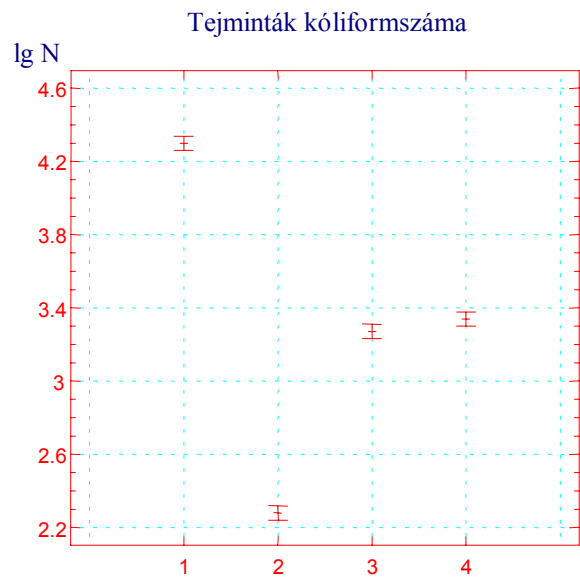
Táptalajokra vonatkozóan  $LSD = 2.06 \cdot 0.080 \cdot \sqrt{\frac{2}{4-3}} = 0.067$  lg egység

Tejmintákra vonatkozóan  $LSD = 2.06 \cdot 0.080 \cdot \sqrt{\frac{2}{3-3}} = 0.078$  lg egység

Az átlagértékek közötti eltéréseket a 4./2. és 4./3. ábrák szemléltetik



**4./2. ábra.**  
Táptalajok átlagértékei



**4./3. ábra.**  
Tejminták kóliformszáma



## 5. LINEÁRIS REGRESSZIÓ

Kvantitatív változók sztochasztikus kapcsolatának vizsgálatára leggyakrabban alkalmazott módszer a regresszió-számítás. A regresszió-számítás során matematikai összefüggést keresünk a változók között. Az  $y = f(x_1, x_2, \dots, x_n)$  összefüggésben a függő változó ( $y$ ) valószínűségi változó, míg a független változók ( $x_1, x_2, \dots, x_n$ ) a leggyakoribb esetben nem valószínűségi változók. Például amikor a mikroorganizmusok, vagy növények szaporodási sebességét vizsgáljuk a táptalaj szén-, nitrogén-, foszfor-tartalmának függvényében, a független változónak tekintett szaporodási sebesség valószínűségi változó (véletlen hatások is befolyásolják), míg az általunk beállított C-, N- és P-koncentrációk nem valószínűségi változók.

Jegyzetünkben a leggyakrabban használt egy független változós lineáris regresszióval foglalkozunk, melynek általános alakja  $y = m \cdot x + a$ , egy egyenes egyenlete, melynek meredeksége  $m$ , tengelymetszete pedig  $a$ . A továbbiakban a független változó ( $x$ ) értékeit nem tekintjük valószínűségi változónak.

Nagyon sok esetben a nem lineáris kapcsolatok is linearizálhatók a változók megfelelő transzformációja révén. Ilyenkor azonban előfordulhat, hogy a transzformáció eredményeképpen kapott új változók nem felelnek meg a lineáris regresszió feltételeinek. A regresszió-számítás feltételei:

- Az  $x$  független változó bármely értékénél az  $y$  valószínűségi változó normális eloszlású legyen
- Az  $y$  valószínűségi változó szórása független legyen az  $x$  változótól.

A mikrobiológiai gyakorlatban mind a normális eloszlás, mind pedig a szórás-homogenitás az alapadatok logaritmikus transzformációjával biztosítható. Minden olyan regressziós feladatnál, ahol a sejtszám változását vizsgáljuk valamilyen független változó függvényében, a lineáris regressziós összefüggést a sejtszám logaritmusára ( $\lg N$ ) vonatkozóan határozzuk meg.

A lineáris regresszió (Gauss nevéhez fűződő) matematikai apparátusa a legkisebb négyzetek elvén alapszik. Az összefüggés  $m$  és  $a$  paramétereit úgy határozzuk meg, hogy a mért és az összefüggésből számított  $y$  értékek eltérésének négyzete minimális legyen:

$$\sum_{i=1}^n [y_i - (a + m \cdot x_i)]^2 \rightarrow \text{minimum}$$

A minimalizálási feladat a fenti célfüggvény  $a$  és  $m$  szerinti parciális deriváltjainak 0-val való egyenlővé tétele után megoldható.

$$m = \frac{\sum_{i=1}^n x_i \cdot y_i - \frac{\sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n}}{\sum_{i=1}^n x_i^2 - \frac{\left[ \sum_{i=1}^n x_i \right]^2}{n}} \quad a = \frac{\sum_{i=1}^n y_i}{n} - b \cdot \frac{\sum_{i=1}^n x_i}{n}$$

ahol  $n$  az összetartozó  $x_i, y_i$  adatpárok száma.

A továbbiakban Sváb (1973) jelölésrendszerét alkalmazva, a regressziószámításokat az alábbi négyzetösszegek felhasználásával végezzük.

$$SQ_x = \sum_{i=1}^n x_i^2 - \frac{\left[ \sum_{i=1}^n x_i \right]^2}{n}, \quad SQ_y = \sum_{i=1}^n y_i^2 - \frac{\left[ \sum_{i=1}^n y_i \right]^2}{n}$$

$$SP = \sum_{i=1}^n x_i \cdot y_i - \frac{\sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n}, \quad SQ_H = SQ_y - \frac{SP^2}{SQ_x}$$

Átlagértékek:  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$

Az egyenes egyenletének együtthatói: meredekség  $m = \frac{SP}{SQ_x}$

tengelymetszet  $a = \bar{y} - m \cdot \bar{x}$

A mérési pontok és a számított egyenes közötti illeszkedés szorosságát a determinációs együttható adja meg. A determinációs együttható ( $R^2$ ) kifejezi, hogy a függő változó (y) értékei közötti eltérés hányadrésze vezethető vissza a lineáris összefüggésre,  $1-R^2$  pedig a véletlenre visszavezethető eltérések arányát adja meg.

$$R^2 = \frac{SP^2}{SQ_x \cdot SQ_y}$$

A gyakorlatban a determinációs együttható helyett nagyon sokszor annak előjeles négyzetgyökét, a korrelációs együtthatót szokták használni a kapcsolat szorosságának jellemzésére. Ennek előnye, hogy előjele révén a kapcsolat irányára is utal.

$$R = \frac{SP}{\sqrt{SQ_x \cdot SQ_y}}$$

A korrelációs együttható értéke (hasonlóan a determinációs együtthatóhoz) 0 és 1 között változik. A 0-hoz közeleső értékek gyenge, vagy nem szignifikáns kapcsolatra utalnak, míg teljes függvénykapcsolat esetén  $R=1$ . A korrelációs együttható szignifikáns kapcsolatra utaló küszöb-értékeit a regresszió-számítás szabadsági foka ( $n-2$ ) függvényében a Függelék 7. táblázata tartalmazza.

A regressziós egyenes körüli szórás:

$$S_{yx}^2 = \frac{1}{n-2} \left[ SQ_y - \frac{SP^2}{SQ_x} \right] \quad S_{yx} = \sqrt{S_{yx}^2}$$

A meredekség (regressziós együttható, iránytangens) hibája:

$$S_m^2 = \frac{S_{yx}^2}{SQ_x} \quad S_m = \sqrt{S_m^2}$$

A tengelymetszet (regressziós állandó) hibája:

$$S_a^2 = S_{yx}^2 \cdot \left[ \frac{1}{n} + \frac{\bar{x}^2}{SQ_x} \right] \quad S_a = \sqrt{S_a^2}$$

A meredekség és a tengelymetszet hibájának ismeretében eldönthető, hogy  $m$  és  $a$  értékei szignifikánsan különböznek-e egy hipotetikus  $m'$  és  $a'$  értéktől.

Próbastatisztika:

$$t_m = \frac{m - m'}{S_m}, \quad t_a = \frac{a - a'}{S_a}$$

Ha a számított értékek nagyobbak, mint az  $n-2$  szabadsági fokhoz tartozó táblázatos  $t$  érték, az összehasonlított paraméterek között az adott szinten szignifikáns különbség van.

Nagyon sokszor felvetődik az a kérdés, hogy van-e egyáltalán összefüggés a függő és független változó között, azaz  $y$  értéke változik-e  $x$  függvényében. Amennyiben nincs összefüggés, akkor  $y$  független  $x$ -től, azaz a számított egyenes meredeksége nem különbözik szignifikánsan 0-tól. Ennek eldöntése érdekében a próbastatisztikában  $m'=0$ , és  $t_m$  értéke alapján dönthetünk arról, hogy  $m$  értéke szignifikánsan eltér-e 0-tól. (Hasonló próba természetesen a tengelymetszetre vonatkozóan is elvégezhető, de ez nincs kapcsolatban az összefüggés szorosságával.)

Ha a számított meredekséget egy másik, szintén regresszióval számított meredekséghez viszonyítjuk, akkor a próbastatisztika számítása:

$$t = \frac{m_1 - m_2}{S_d}, \text{ ahol} \quad S_d = \sqrt{\frac{SQ_{H1} + SQ_{H2}}{n_1 + n_2 - 4} \left[ \frac{1}{SQ_{x1}} + \frac{1}{SQ_{x2}} \right]}$$

Ha  $S_{yx1}$  szignifikánsan különbözik  $S_{yx2}$ -től, akkor  $S_d = \sqrt{S_{m1}^2 + S_{m2}^2}$

Annak eldöntésére, hogy az egyenes egyenlete jól írja-e le az összefüggést, illetve egy más függvény alkalmazása esetén nem szorosabb-e az illeszkedés, variancia-analízis is végezhető. Ekkor az összehasonlítandó függvényekből számított véletlen szórásnégyzeteket ( $S_{yx}^2$ ) hasonlítjuk össze, és F-próbával döntünk.

A regressziós összefüggés konfidencia-intervallumának ( $\pm h$ ) számítása minden egyes  $x$  értékhez külön történik:

$$y = m \cdot x + a \pm h$$

Egyedi értékekre: 
$$h_y = t \cdot \sqrt{S_{yx}^2 \left[ 1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{SQ_x} \right]}$$

Átlagértékekre: 
$$h_{\bar{y}} = t \cdot \sqrt{S_{yx}^2 \left[ \frac{1}{n} + \frac{(x - \bar{x})^2}{SQ_x} \right]}$$
 A  $t$  érték szabadsági foka  $n-2$ .

A számított értékek konfidencia-intervallumában szereplő  $(x - \bar{x})^2$  tag miatt a konfidencia-intervallumnak  $\bar{x}$  értéknél minimuma van. Ettől távolodva a konfidencia-intervallum szélesedik.

### 5.1. Példa

Hőkezelés méretezéshez egy adott mikrobapopuláció túlélését vizsgálták az idő függvényében. A vizsgálatok célja az adott hőmérséklethez tartozó tizedelődési idő meghatározása volt. A túlélő sejtszám logaritmus értékeit az 5./1. Táblázatban foglaltuk össze.

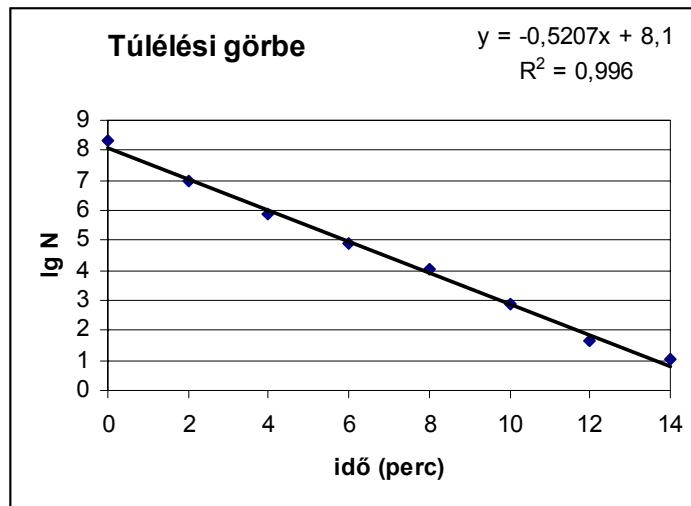
#### 5.1. Táblázat. Hőkezelést túlélő sejtek száma (lg N/ml)

idő (perc)	lg N
0	8.30
2	7.00
4	5.86
6	4.90
8	4.04
10	2.86
12	1.64
14	1.04

Az összefüggést a STATGRAPHIC 5.1. programmal kiértékelve, az eredményeket az 5./2 és 5./3. táblázatokban foglaltuk össze, valamint az 5./1. ábrával szemléltetjük.

#### 5.2. Táblázat. Regressziószámítás variancia-táblázata

Variancia forrás	Négyzetössze g	Sz.. fok	S <sup>2</sup>	F	Szign. szint
Összes	45.7358	7			
Model	45.5521	1	45.5	1487	0.0000
Véletlen	0.1837	6	0.0306 = S <sup>2</sup> <sub>yx</sub>		



**5./1.** ábra. Hőpusztulási kísérlet túlélési görbéje

**5./3. Táblázat.** Regressziós egyenes paraméterei

Paraméter		Standard error	t	Sign. szint
Tengelymetszet	a = 8.10	0.113	71.7	0.0000
Meredekség	m = - 0.5207	0.0135	-38.57	0.0000
Determinációs együttható	$R^2 = 0.9960$	Korrelációs együttható	$R = -0.9980$	

A regressziós egyenlet paramétereinek konfidencia-intervallum számításához a t érték szabadsági foka megegyezik a regresszió-számítás véletlen hibájának szabadsági fokával, esetünkben  $Szf=6$ .

Kétoldali  $\alpha=5\%$  elsőfajú hibával  $t = 2.45$

*Tengelymetszet konfidencia-intervalluma:*  $\Delta_a = 2.45 \cdot 0.113 = 0.277$        $a = 8.10 \pm 0.28$

*Meredekség konfidencia-intervalluma:*  $\Delta_m = 2.45 \cdot 0.0135 = 0.033$        $m = -0.521 \pm 0.033$

A túlélési görbe meredekségéből számított tizedelőési idő:       **$D = -2.303/m = 4.42$  perc**

A származtatott mennyiségek konfidencia-intervallumának meghatározásakor óvatosan kell eljárunk. Esetünkben a meredekség konfidencia-intervallumának két szélső értékéből (-0.554 és -0.488) határozzuk meg a tizedelőési idő 95%-oskonfidenciaintervallumát:

**$4.16 \leq D < 4.72$  perc**

## 6. ÉLŐSEJTSZÁM MEGHATÁROZÁSI MÓDSZEREK HIBÁJA

A tenyésztéses élősejtszám meghatározási módszerek hibája a vizsgálati anyag biológiai jellegénél fogva jelentősen nagyobb, mint a fiziko-kémiai analitikai eljárásoké. Ennek ellenére azonban a kísérleti hiba meghatározására teljesen ugyanazok a matematikai-statisztikai módszerek szolgálnak, a különbség csak annyi, hogy a relatív szórás esetünkben nagyságrendekkel nagyobb lehet, mint a klasszikus analitikai módszerek esetében.

Tenyésztéses élősejtszám meghatározásra alapvetően két módszer terjedt el a mikrobiológiai gyakorlatban: határhígítási és lemezöntéses módszer. A két módszernek sokféle változata és továbbfejlesztése létezik, azonban ezek az alapeljárástól lényegükben nem különböznek.

### 6.1. HATÁRHÍGÍTÁSOS (MPN) MÓDSZER

A határhígítási módszer lényege, hogy a vizsgálandó anyagból alapszuspenziót készítünk, amelyet addig hígítunk, amíg az utolsó hígítási tagokban nincs mikroba. Az utolsó hígításokból párhuzamos leoltásokat végzünk táptalajba és vizsgáljuk a szaporodást. A szaporodást mutató párhuzamos csövek hígítási szintek szerinti megoszlása alapján matematikai-statisztikai alapon következtetünk az eredeti szuszpenzió legvalószínűbb élősejtszámára (Most Probable Number, MPN). A leggyakrabban alkalmazott módszer szerint hígításonként 3 párhuzamos leoltást végeznek. A kulcsszámok meghatározásához olyan hígítási szinteket választanak, amelyek lehetőleg maximális számú (3) pozitív csővel kezdődnek és ezután a pozitív csövek száma kevesebb. Pl. a 310 kulcsszám azt jelenti, hogy a figyelembe vett hígítások első tagjában 3, a másodikban 1, a harmadikban pedig 0 pozitív cső van. Az un. Hoskins táblázatból kikereshető a 310 kulcsszámhoz tartozó MPN érték (4.3), ezt megszorozva az első figyelembe vett hígítási szinttel, kapjuk az eredeti szuszpenzió mikrobaszámát.

Az eljárás a következő matematikai megfontolásokon alapszik.

$s_i$  db negatív (steril) cső előfordulásának a valószínűsége az  $i$ -edik hígításban:

$$P_i = \frac{n!}{s_i! \cdot (n - s_i)!} \cdot p_i^{s_i} \cdot (1 - p_i)^{n - s_i},$$

ahol:  $P_i$  annak a valószínűsége, hogy  $n$  db. beoltott csőből  $s_i$  db steril.  
 $p_i$  a negatív cső valószínűsége az adott hígításban

A mikrobaszámok Poisson eloszlása következtében  $p_i = e^{-\lambda_i}$ , ahol  $\lambda_i$  az inokulumban lévő mikrobák száma.

Tízszeres hígításokat figyelembe véve, három egymást követő hígítási tag mikrobaszáma:

$$\begin{array}{ll} \lambda_1 = N & p_1 = e^{-N} \\ \lambda_2 = 0.1 \cdot N & p_2 = e^{-0.1 \cdot N} \\ \lambda_3 = 0.01 \cdot N & p_3 = e^{-0.01 \cdot N} \end{array}$$

Hígításonként három párhuzamos leoltással, a steril csövek valószínűsége:

$$P_1 = \frac{3!}{s_1! \cdot (3 - s_1)!} \cdot p_1^{s_1} \cdot (1 - p_1)^{3 - s_1}$$

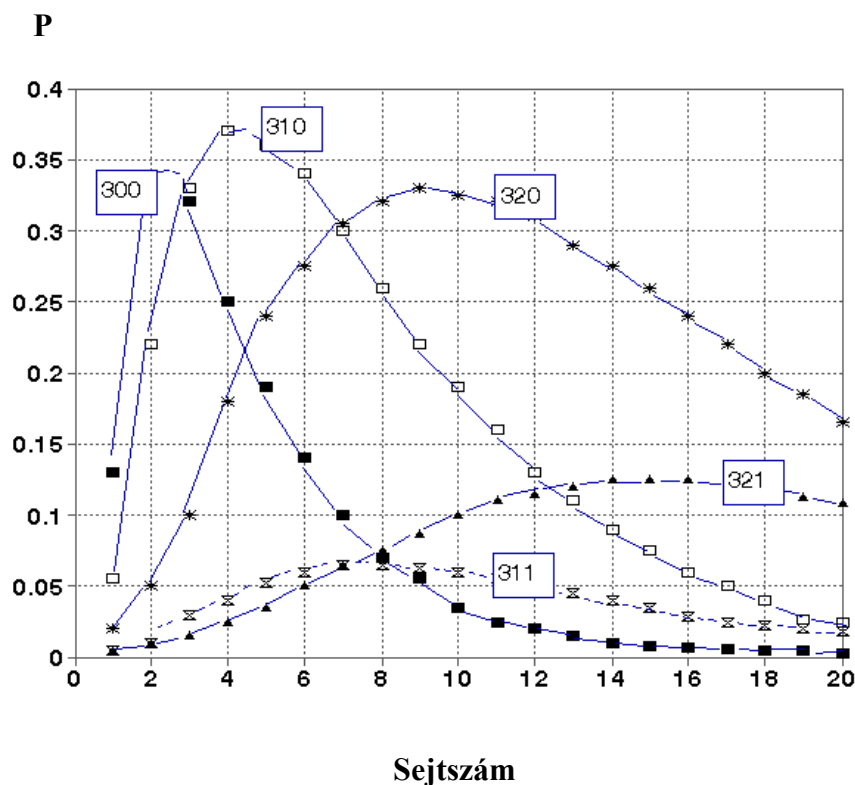
$$P_2 = \frac{3!}{s_2! \cdot (3 - s_2)!} \cdot p_2^{s_2} \cdot (1 - p_2)^{3 - s_2}$$

$$P_3 = \frac{3!}{s_3! \cdot (3 - s_3)!} \cdot p_3^{s_3} \cdot (1 - p_3)^{3 - s_3}$$

A három egymást követő hígítási szinten az  $s_1, s_2, s_3$  db steril cső együttes előfordulásának valószínűsége:

$$P = P_1 \cdot P_2 \cdot P_3$$

Kiszámítva a különböző sejtszámokhoz tartozó  $P$  valószínűséget, bármely  $s_1, s_2, s_3$  kombinációhoz (kivéve a 0, 0, 0 és 3, 3, 3 kombinációkat), meghatározható a valószínűség eloszlás sűrűségi görbéje. A konvencióknak megfelelően a valószínűségeket nem a negatív (steril), hanem a pozitív (szaporodást mutató) csövek kombinációjának függvényében határozták meg. A valószínűségi görbék maximumához tartozó sejtszámok adják az adott kombinációnak megfelelő legvalószínűbb sejtszámot (MPN), amelyet táblázatosan szoktak közölni. Néhány jellegzetes valószínűségi görbét mutatunk be a 6./1. ábrán. A 3 hígítási szinthez, szintenként 3 párhuzamos leoltáshoz tartozó MPN értékeket és a hozzájuk tartozó  $P_{\max}$  maximális valószínűségeket a 6./1. táblázat tartalmazza.



**6./1. ábra.** Kulcsszámok valószínűsége a sejtszám függvényében

**6./1. Táblázat**

MPN táblázat tízszeres hígítás, 3 párhuzamos leoltás

<b>kulcsszám</b>	<b>P<sub>max</sub></b>	<b>MPN</b>	<b>Kulcsszám</b>	<b>P<sub>max</sub></b>	<b>MPN</b>
000		<0.3	<b>200 *</b>	<b>0.3193</b>	<b>0.91</b>
001	0.0033	0.3	201 *	0.0112	1.4
002	1.5·10 <sup>-5</sup>	0.6	202	1.9·10 <sup>-4</sup>	2.0
003	3.7·10 <sup>-8</sup>	0.9	203	1.4·10 <sup>-6</sup>	12.6
010 **	0.0336	0.3	<b>210 *</b>	<b>0.1196</b>	<b>1.5</b>
011	4.5·10 <sup>-4</sup>	0.61	211 *	0.0063	2.0
012	3.5·10 <sup>-6</sup>	0.92	212	1.5·10 <sup>-4</sup>	2.7
013	1.2·10 <sup>-8</sup>	1.2	213	1.5·10 <sup>-6</sup>	3.4
020	0.0016	0.62	220 *	0.0232	2.1
021	3.6·10 <sup>-5</sup>	0.93	221	0.0017	2.8
022	3.9·10 <sup>-7</sup>	1.2	222	5.4·10 <sup>-5</sup>	3.5
023	1.8·10 <sup>-9</sup>	1.6	223	7.0·10 <sup>-7</sup>	4.2
030	4.2·10 <sup>-5</sup>	0.94	230	0.0022	2.9
031	1.4·10 <sup>-6</sup>	1.3	231	2.1·10 <sup>-4</sup>	3.6
032	2.0·10 <sup>-8</sup>	1.6	232	8.6·10 <sup>-6</sup>	4.4
033	1.2·10 <sup>-10</sup>	1.9	233	1.4·10 <sup>-7</sup>	5.3
<b>100 *</b>	<b>0.3920</b>	<b>0.36</b>	<b>300 *</b>	<b>0.3410</b>	<b>2.3</b>
101 **	0.0062	0.72	301 *	0.0310	3.9
102	5.6·10 <sup>-5</sup>	1.1	302	0.0016	6.4
103	2.4·10 <sup>-7</sup>	1.5	303	4.3·10 <sup>-5</sup>	9.5
110 *	0.0645	0.73	<b>310 *</b>	<b>0.3743</b>	<b>4.3</b>
111	0.0018	1.1	311 *	0.0658	7.5
112	2.4·10 <sup>-5</sup>	1.5	312	0.0065	12
113	1.4·10 <sup>-7</sup>	1.9	313	3.2·10 <sup>-4</sup>	16
120 **	0.0063	1.1	<b>320 *</b>	<b>0.3282</b>	<b>9.3</b>
121	2.5·10 <sup>-4</sup>	1.5	<b>321 *</b>	<b>0.1251</b>	<b>15</b>
122	4.4·10 <sup>-6</sup>	2.0	322 **	0.0247	21
123	3.2·10 <sup>-8</sup>	2.4	323	0.0024	29
130	3.0·10 <sup>-4</sup>	1.6	<b>330 *</b>	<b>0.3659</b>	<b>24</b>
131	1.6·10 <sup>-5</sup>	2.0	<b>331 *</b>	<b>0.4277</b>	<b>46</b>
132	3.6·10 <sup>-7</sup>	2.4	<b>332 *</b>	<b>0.4444</b>	<b>110</b>
133	3.2·10 <sup>-9</sup>	2.9	333		>110

\* 1. kategória. Az esetek 95%-ában előfordul

\*\* 2. kategória. Az eseteknek csak 4%-ában fordul elő.

A 6./1. táblázatban feltüntetett P<sub>max</sub> értékek alapján egyértelmű, hogy az elméletileg lehetséges kulcsszámok közül csak néhány, a táblázatban vastagon jelölt kombináció bír gyakorlati jelentőséggel. Ezek a következők: 100, 200, 210, 300, 310, 320 és 321.

A 330, 331 és 332 gyakorlatilag megegyezik az eggyel nagyobb hígításból származó 300, 310 és 320 kulcsszámokkal. Minden más kombináció igen nagy valószínűséggel befertőződés, vagy helytelen hígítás következménye.

Az MPN értékek konfidencia-intervallumának meghatározásához szorzófaktorokat használunk, melyeket a hígítási lépték és a hígításonkénti párhuzamos leoltások számának



függvényében a 6./2. Táblázatban foglaltunk össze. A szorzó-faktorok a hígításból eredő hibát nem tartalmazzák.

### **6./2. Táblázat.**

Az MPN értékek 95%-os konfidencia-intervallumának szorzó-faktorai.

Párhuzamos leoltások száma hígításonként	Hígítási lépték			
	2	4	5	10
1	4.0	7.14	8.32	14.45
2	2.67	4.00	4.47	6.61
3	2.23	3.10	3.39	4.68
4	2.00	2.68	2.88	3.80
5	1.86	2.41	2.58	3.30
10	1.55	1.86	1.95	2.32

### **6.1. Példa**

Tejminta coliform számának meghatározását végezzük határhígításos módszerrel. Tízes hígítási léptéket és hígításonként 3-3 párhuzamosban 1-1 ml-nyi leoltást alkalmazva, az alábbi eredményeket kaptuk.

Hígítási szint	Coliform
1.	+++
2.	+++
3.	+--
4.	---
5.	---

A 2., 3. és 4. hígítás eredményeit figyelembe véve a kulcsszám: 3 1 0.

A 310 kulcsszámnak megfelelő MPN érték a 6./1. táblázatból kikeresve: MPN = 4.3.

Az eredeti minta coliform számának legvalószínűbb értéke:  $4.3 \cdot 10^2 / \text{ml}$ .

A meghatározott coliform szám konfidencia-intervallumának kiszámításához a 6./2. Táblázatból keressük ki a szorzó-faktort. Ennek értéke 10-es hígítási lépték és hígításonként 3 párhuzamos leoltás esetén 4.68.

A konfidencia-intervallum alsó és felső határát megkapjuk, ha az MPN értéket osztjuk és szorozzuk a megfelelő faktorial, melynek értéke esetünkben 4.68.

Alsó határ:  $4.3 / 4.68 = 0.92$

Felső határ:  $4.3 \cdot 4.68 = 20.12$

A tejminta coliform száma 95%-os biztonsággal az alábbi sávba esik:

$$9.2 \cdot 10^1 \leq 4.3 \cdot 10^2 < 2.0 \cdot 10^3$$

### Kísérlettervezési szempontok

A szokásos mikrobiológiai gyakorlatban alkalmazott tízes hígítási lépték és hígításonként 3 párhuzamos leoltás esetén a konfidencia-intervallum számításához használt 4.68-as szorzófaktor egy nagyságrendnél szélesebb ( $4.68^2 = 21.9$ ) konfidencia-sávot eredményez. A széles konfidencia-intervallum miatt az MPN módszer alkalmazása csak speciális esetekben indokolt. További problémát jelenthet, hogy az MPN értékek kvantáltak, gyakorlatilag csak 4–5 számértéknek van jelentősége, ezért nem kezelhető folytonos valószínűségi változóként. Az adatok matematikai-statisztikai feldolgozásakor ez elméleti problémákat okozhat.

*Az MPN módszer alkalmazásának indokai:*

- Igényes mikroorganizmusok szaporodása leves-tenyészetben jobb, mint agaron.
- Kicsi (10 alatti) sejtszámok biztonságos kimutatása csak MPN módszerrel oldható meg.
- Leves-tenyészeteket alkalmazva lehetőségünk van a gázképződés indikálására és egyes biokémiai reakciók közvetlen elvégzésére.

*Nem célszerű MPN módszert alkalmazni az alábbi esetekben:*

- Egy nagyságrendnél kisebb mikrobaszám-változások nyomonkövetése.
- Nagyságrenden belüli adatok folytonos valószínűségi változót igénylő matematikai-statisztikai kiértékeléséhez MPN alapadatok nem használhatók.
- Amikor fennáll a csövek befertőződésének lehetősége.

## **6.2. LEMEZÖNTÉSES ÉS FELÜLETI SZÉLESZTÉSES MÓDSZER**

A vizsgálandó anyagból alapsuszpenziót készítünk, majd hígítási sort. A hígítási sor tagjaiból lemezt öntünk, vagy előre kiöntött és megszilárdult lemezek felületére szélesztünk. A szükséges inkubálás után a kifejlődött telepeket megszámloljuk, majd a telepszámok és a hígítási szint ismeretében kiszámítjuk a minta sejtszámát.

A minta sejtszámának (N) számítása a következő:

$$N = H_0 \cdot (V/v)^h \cdot N_t / v_i$$

ahol  $H_0$ : az alapsuszpenzió hígítási szintje (általában 10)  
 $V/v$ : a hígítási lépték ( $V$  = hígítóoldat + inokulum,  $v$  = inokulum,  $V/v$  általában 10)  
 $h$ : a hígítási szintek száma (tetszőlegesen változtatható)  
 $N_t$ : A  $h$ -adik hígítási szintnél leolvasott telepszám  
 $v_i$ : A Petri csészére felvitt inokulum mennyisége (lemezőntésnél általában 1.0 ml, szélesztésnél 0.1 ml)

Ha az alapsuszpenzió is 10-szeres hígítású, úgy ezt tekintjük a hígítási sor első tagjának.

A módszer hibaforrásai az alábbiak.

- Alapsuszpenzió készítésének hibája
- Hígítási hiba
- Lemezre-vitel hibája
- Sejtszám-eloszlás hibája

*Az alapsuszpenzió készítésénél* (amennyiben erre szükség van) általában 10 g-ot mérünk be 90 ml hígító folyadékba. Ennek a bemérési eljárásnak a hibája a többi lépéshez képest elhanyagolható.

*A hígítási hiba* a pipettázott térfogatok (általában 1.0 ml) hibájának halmozódásából áll.

- A hígító-pipetták ( $v=1.0$  ml) átlagos szórása:  $\sigma_v = 0.03$ ml,  $\sigma_v/v = 0.03$ .
- A hígító közeg térfogatának ( $V = 10$  ml) átlagos szórása:  $\sigma_V = 0.3$  ml.  $\sigma_V/V = 0.03$
- *A lemezre vitel hibája* a pipettahibából adódik:  $\sigma_{v_i} = 0.03$  ml.  $\sigma_{v_i}/v = 0.03$ .
- *A sejtszám-eloszlás hibája* a sejtszámok Poisson-eloszlásából következően:  $\sigma_N = \sqrt{N}$

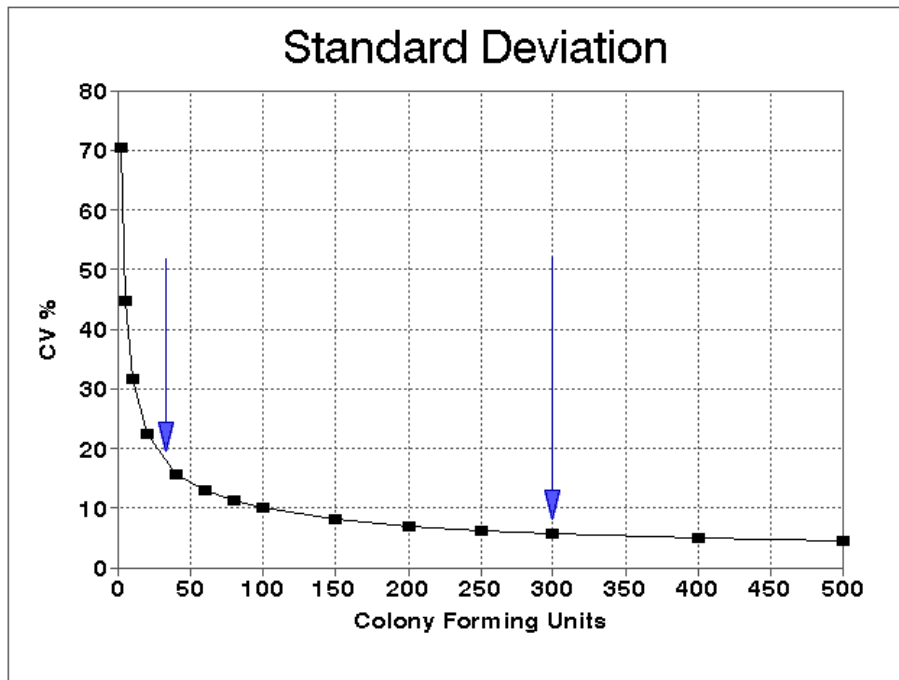
A telepszámok (colony forming units, CFU) szórása (SD), relatív szórása (CV%) és 95%-os konfidencia-intervalluma a 6./3. Táblázatban található.

**6./3. Táblázat.** Telepszámok relatív szórása és konfidencia-intervalluma

CFU	Standard Deviation		Konfidencia-intervallum	
	SD = $\sqrt{N}$	CV (%)	2·SD	(%)
<b>2</b>	1.41	70.5	± 2.8 *	± 141 *
<b>5</b>	2.24	44.8	± 4.48 *	± 89.6 *
<b>10</b>	3.16	31.6	± 6.32 *	± 63.2 *
<b>20</b>	4.47	22.4	± 8.94	± 44.8
<b>40</b>	6.32	15.8	± 12.6	± 31.6
<b>60</b>	7.75	12.9	± 15.5	± 25.8
<b>80</b>	8.94	11.2	± 17.9	± 22.4
<b>100</b>	10.00	10.0	± 20.0	± 20.0
<b>150</b>	12.25	8.17	± 24.5	± 16.3
<b>200</b>	14.14	7.07	± 28.3	± 14.1
<b>250</b>	15.81	6.32	± 31.6	± 12.6
<b>300</b>	17.32	5.77	± 34.6	± 11.5
<b>400</b>	20.00	5.00	± 40.0	± 10.0
<b>500</b>	22.36	4.47	± 44.7	± 8.9
<b>1000</b>	31.62	3.62	± 63.2	± 6.3

\*: Csak tájékoztató érték. Korrekt módon Poisson-eloszlásból számítható, aszimmetrikus konfidencia-intervallum.

A százalékban kifejezett szórásértékeket a 6./2. ábra szemlélteti. Az ábrán bejelöltük azt a tartományt, amelyet általában az értékeléshez előírnak. Ez a 30 és 300 közötti telepszámok tartománya. Ez alatt a relatív szórás erősen nő, fölötté pedig a telepek kifejlődése már gátolt és jelentőssé válik a számlálási hiba is.



### 6./2. ábra

Telepszámok relatív szórása

A lemezöntéses módszer eredő hibáját az összes hibaforrás figyelembevételével a hibaátzármaztatás törvényszerűségeinek alapján számíthatjuk ki.

Az eredeti sejtkoncentráció kiszámítására szolgáló összefüggés:  $(N = H_0 \cdot (V/v)^h \cdot N_t/v_t)$  helyett annak logaritmusával számolunk. A logaritmikus transzformáció célja a szórások sejtszámtól való függetlenségének és homogenitásának biztosítása.

$$\lg N = \lg H_0 + h \cdot (\lg V - \lg v) - \lg v_t + \lg N_t$$

Az eredő szórásnégyzet a következő módon számítható:

$$\sigma^2_{\lg N} = \left[ \frac{\partial \lg N}{\partial H_0} \cdot \sigma_{H_0} \right]^2 + \left[ \frac{\partial \lg N}{\partial V} \cdot \sigma_V \right]^2 + \left[ \frac{\partial \lg N}{\partial v} \cdot \sigma_v \right]^2 + \left[ \frac{\partial \lg N}{\partial v_t} \cdot \sigma_{v_t} \right]^2 + \left[ \frac{\partial \lg N}{\partial N_t} \cdot \sigma_{N_t} \right]^2$$

Amint a fentiekben említettük az alapsuszpenzió készítésének hibáját elhanyagoljuk, csak a többi lépéssel foglalkozunk.

Kifejtve a változók szerinti parciális deriváltakat:

$$\sigma_{\lg N}^2 = \left[ \frac{h}{2.303} \cdot \frac{1}{V} \cdot \sigma_V \right]^2 + \left[ \frac{h}{2.303} \cdot \frac{1}{v} \cdot \sigma_v \right]^2 + \left[ -\frac{h}{2.303} \cdot \frac{1}{v_t} \cdot \sigma_{v_t} \right]^2 + \left[ \frac{h}{2.303} \cdot \frac{1}{N_t} \cdot \sigma_{N_t} \right]^2$$

$$\sigma_{\lg N}^2 = \frac{1}{2.303^2} \cdot \left[ h^2 \cdot \left( \frac{\sigma_V^2}{V^2} + \frac{\sigma_v^2}{v^2} \right) + \frac{\sigma_{v_t}^2}{v_t^2} + \frac{\sigma_{N_t}^2}{N_t^2} \right]$$

A fenti összefüggés zárójelében lévő első tag a hígítás, a második a lemezre-vitel, a harmadik pedig a telepszám-eloszlás és számlálás hibáját reprezentálja.

Behelyettesítve az egyenletbe az egyes tagokra jellemző, 82. oldalon megadott 0.03 relatív szórás értéket:

$$\sigma_{\lg N}^2 = \frac{1}{2.303^2} \cdot \left[ h^2 \cdot (0.03^2 + 0.03^2) + 0.03^2 + \frac{\sigma_{N_t}^2}{N_t^2} \right]$$

Kiszámítva az értékeket:

$$\sigma_{\lg N}^2 = \frac{1}{2.303^2} \cdot \left[ h^2 \cdot 1.8 \cdot 10^{-3} + 9 \cdot 10^{-4} + \frac{\sigma_{N_t}^2}{N_t^2} \right]$$

Figyelembe véve, hogy a Poisson eloszlású telepszámnál  $\sigma_N^2 = N$  és a számlálás hibáját 10%-nak véve  $\sigma_{\text{száml.}}^2 = 0,01 \cdot N$ , a telepszám-eloszlás és –számlálás együttes relatív hibája:

$$\frac{\sigma_{N_t}^2}{N_t^2} = \frac{N_t + 0.01 \cdot N_t^2}{N_t^2} = \frac{1}{N_t} + 0.01$$

Behelyettesítve megkapjuk a lemezöntéses eljárással meghatározott  $\lg N$  értékek szórásnégyzetét:

$$\sigma_{\lg N}^2 = \frac{1}{2.303^2} \cdot \left[ h^2 \cdot 1.8 \cdot 10^{-3} + 9 \cdot 10^{-4} + \frac{1}{N_t} + 0.01 \right]$$

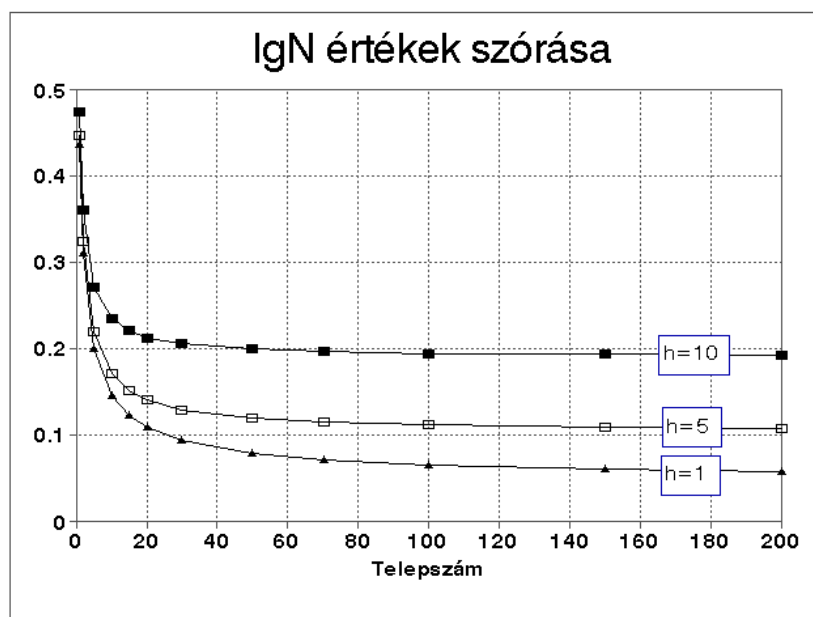
A zárójelen belüli első tag a hígítás, a második a felvitel, a harmadik a telepszám-eloszlás, a negyedik pedig a telepszámlálás hibáját adja meg.

Az összefüggésből jól látható, hogy az eredő szórásnégyzet értéke a hígítással négyzetesen nő, míg a telepszámmal reciprok összefüggésben áll. A felviteli és a számlálási hiba közel állandónak tekinthető, melyen belül a felviteli hiba jelentősen kisebb.

A lemezöntéses módszer eredő hibáját ( $\sigma_{\lg N}$ ) a különböző hígításokban meghatározott telepszámok függvényében a 6./4. Táblázat tartalmazza és a 6./3. ábra szemlélteti.

**6./4. Táblázat.** Lemezöntéses, vagy felületi szélesztéses módszer hibája (h = hígítási szint).

N	$\sigma_{lgN}$		
	h=1	h=5	h=10
1	0.4369	0.4461	0.4738
2	0.3109	0.3237	0.3609
5	0.2002	0.2196	0.2714
10	0.1458	0.1714	0.2342
15	0.1223	0.1520	0.2203
20	0.1087	0.1413	0.2131
30	0.0932	0.1297	0.2056
50	0.0785	0.1196	0.1994
70	0.0713	0.1150	0.1967
100	0.0654	0.1115	0.1946
150	0.0604	0.1086	0.1930
200	0.0578	0.1070	0.1922



**6./3. ábra.**

Lemezöntéses élősejtszám-meghatározás hibája (h = hígítási szint).

### 6.3. HATÁRHÍGÍTÁSOS ÉS LEMEZÖNTÉSES MÓDSZEREK ÖSSZEHA-SONLÍTÁSA

Az elősejtszám-meghatározási módszerek összehasonlító értékelése alapján egyértelműen eldönthető, hogy azok milyen feltételek mellett, milyen sejtszám-tartományban és milyen megbízhatósággal alkalmazhatók. Az összehasonlító értékelést a 6./5. Táblázat tartalmazza.

#### 6./5. Táblázat. Élősejtszám-meghatározási módszerek összehasonlító értékelése.

	<b>Lemezöntés</b>	<b>Határhígítás</b>
<b>Torzítás</b>	Nem torzít.	Esetenként 50% torzítás.
<i>Véletlen hiba elméleti értéke</i>	$\sigma_{lgN} = 0.15$	$\sigma_{lgN} = 0.50$
<i>Egyetlen adat esetén a becslés pontossága</i>	$\Delta = \pm 0.3$	$\Delta = \pm 1$
<i>Hibaforrás</i>	A hígítási hiba a döntő	Az eloszlási hiba a döntő
<i>Hibacsökkentés lehetősége</i>	Párhuzamos hígítási sorok alkalmazása	Párhuzamos leoltások alkalmazása
<i>Célszerű alkalmazási tartomány</i>	$N > 30$	$N < 30$

#### Kísérlettervezési szempontok

Az elősejtszám-meghatározási módszerek megválasztásának szempontjairól a módszerek ismertetésénél már szóltunk. Itt csupán a párhuzamosok megválasztásának fontosságát kívánjuk hangsúlyozni.

Párhuzamos vizsgálatokat a meghatározási módszer azon pontjánál kell beállítani, amelyek a legnagyobb hibaforrást jelentik. Lemezöntéssel eljárásnál a legnagyobb hibaforrást a hígítás jelenti, ezért a kiindulási hígítási sorok számát kell növelni és nem az azonos hígításból való leoltásokat. Célszerű eljárás: két párhuzamos hígítási sor, hígításonként 1-1 lemez öntésével. A különböző hígításokból figyelembe vett telepszámokból az eredeti szuszpenzió mikrobaszáma az alábbi összefüggéssel számítható:

#### **6.2. Példa**

Lemezöntéssel eljárás során egyetlen hígítási sorból az alábbi eredményeket kaptuk: 3. hígításban 320 telepszám, 4. hígításban 28 telepszám. Mindkét telepszám a jól értékelhető tartományba esik, ezért mind a kettőt figyelembe vesszük. Az eredeti szuszpenzió mikrobaszáma (N) az alábbi összefüggéssel számítható:

$$N = \frac{\sum N_i}{v_i \cdot (n_1 + 0.1 \cdot n_2) \cdot d}$$

ahol  $N_i$  az összes leszámolt telep,  
 $n_1$  a kisebbik hígításban figyelembevett Petri csészék száma, esetünkben 1.  
 $n_2$  a következő hígításban figyelembe vett Petri csészék száma, szintén 1.  
 $v_i$  az inokulum mennyisége, esetünkben 1 ml.  
 $d$  a kisebbik hígítási szint hígítási faktora, a 3. hígításban  $10^{-3}$ .

Behelyettesítve a kapott telepszámokat: 
$$N = \frac{320 + 28}{1 \cdot (1 + 0.1 \cdot 1) \cdot 10^{-3}} = \frac{348}{1.1 \cdot 10^{-3}} = 3.16 \cdot 10^5 / \text{ml}.$$

#### 6.4. MÓDSZER ISMÉTELHETŐSÉGÉNEK ÉS REPRODUKÁLHATÓSÁGÁNAK MEGHATÁROZÁSA

Egy analitikai eljárás mérési bizonytalanságának két legfontosabb jellemzője az ismételhetőség és a reprodukálhatóság.

**Ismételhetőség** alatt általában az azonos laboratóriumon belül, azonos eszközökkel, azonos személy által végzett meghatározások eredményei közötti egyezés mértékét értjük. Jellemzésére a párhuzamos vizsgálatok szórása ( $s_r$ ) szolgál, és értéke a véletlen hatásokat tükrözi.

**Reprodukálhatóság** alatt általában a különböző laboratóriumok által, de laboratóriumon belül az ismételhetőségnek megfelelően végzett meghatározások eredményei közötti egyezés mértékét értjük. A reprodukálhatóságra jellemző szórás ( $s_R$ ) a véletlen hatásokra ( $s_r$ ) és a laboratóriumok közötti eltérésekre visszavezethető ( $s_L$ ) szórások eredőjeként határozható meg:

$$s_R^2 = s_L^2 + s_r^2$$

Az analitikai teljesítményjellemzők esetében néha megkülönböztetik a ***laboratóriumon belüli reprodukálhatóságot*** is, melyen általában az azonos laboratóriumon belül eltérő eszközökkel, személyekkel végzett vizsgálatok eredményei közötti egyezés mértékét értik. Nyilvánvaló, hogy ebben az esetben a véletlen hatások mellett az eszközök és a személyek közötti eltérésekből eredő hatások is érvényesülnek.

Egy módszer ismételhetősége egyetlen laboratóriumon belül is meghatározható párhuzamos mérésekkel, ez az érték azonban csak az adott laboratórium eredményeire lesz jellemző. Más laboratóriumban esetleg kisebb, vagy nagyobb szórással dolgozhatnak.

Egy analitikai eljárás reprodukálhatóságának meghatározása csak laboratóriumok közötti módszer összehasonlító vizsgálatok alapján végezhető el. Korrekt kivitelezése gondos kísérlettervezést és általában variancia-analízissel való kiértékelést igényel.

##### **Kísérlettervezési szempontok**

A laboratóriumok közötti módszer összehasonlító vizsgálat célja a véletlen hiba átlagos értékének meghatározása mellett a laboratóriumok közötti eltérésekre jellemző hiba meghatározása. A kétféle hiba (illetve szórásérték) felhasználásával a módszerre jellemző reprodukálhatóság kiszámítható. A kiértékelés egytényezős variancia-analízissel végezhető el, ezért az alapadatoknak ki kell elégíteniük a variancia-analízis feltételeit:

- Az eredmények a kezelésektől eltekintve függetlenek legyenek egymástól.
- Az egyes kezeléseken belül az eredmények eloszlása normális legyen.
- A kezeléseken belüli véletlen szórásnégyzetek között ne legyen szignifikáns különbség.

A véletlen hiba meghatározásánál alkalmazandó párhuzamosok számánál vegyük figyelembe, hogy a véletlen szórásnégyzet szabadsági foka  $k \cdot (p-1)$ , ahol esetünkben  $k$  a laboratóriumok,  $p$  pedig a laboratóriumon belüli párhuzamosok száma. Mivel az F-próba érzékenysége a nevező 10 feletti szabadsági fokainál lecsökken, 5-nél több kezelés esetén a kezeléseken belüli párhuzamosok számát nem érdemes 3-nál nagyobbra választani.



### 6. 3. Példa

#### Lemezöntéses módszer ismételtetésének és reprodukálhatóságának meghatározása pasztörtej összes mikrobaszámára vonatkozóan

A 20 laboratórium részvételével szervezett módszer összehasonlító vizsgálatához egy referencia laboratórium által kiadott pasztörtej mintákat használtunk fel. A vizsgálatok céljára a minták csak abban az esetben használhatók, ha azok homogenitásáról előzetesen meggyőződünk. Homogenitás alatt ebben az esetben azt értjük, hogy nincs közöttük kiugróan nagy vagy kis érték, amely a feltételezett normális eloszlástól eltérne.

A variancia-analízis feltételeit az alábbiak szerint biztosítottuk.

- Az alapadatok normális eloszlása érdekében a milliliterenkénti sejtszámok logaritmusát használtuk a számításokhoz.
- A kiugró értékek vizsgálatához Grubbs tesztet alkalmaztunk.
- A szórások homogenitását (kiugróan nagy szórások vizsgálatát) Cochran próbával ellenőriztük.

#### **Kiadott minták homogenitásának vizsgálata**

A vizsgálatokra készített mintákból 8 mintaelemet megvizsgálva, azok alapadatait a 6./6. táblázatban foglaltuk össze.

#### **6./6. Táblázat.**

Pasztörtej minták összes mikrobaszáma. N (cfu/ml)

	1.	2.	3.	4.	5.	6.	7.	8.
<b>N</b>	$3.2 \cdot 10^4$	$4.6 \cdot 10^4$	$3.8 \cdot 10^4$	$5.2 \cdot 10^4$	$2.6 \cdot 10^4$	$4.5 \cdot 10^4$	$2.4 \cdot 10^4$	$3.7 \cdot 10^4$
<b>lg N</b>	4.51	4.66	4.58	4.72	4.41	4.65	4.38	4.57

A kiugró értékek próbáját a lg N értékekre végezzük el.

Lehetőségünk van Dixon- vagy Grubbs-próbát alkalmazni.

A Dixon-próba a Függelék 9. táblázat alapján elvégezhető, az alábbiakban a Grubbs próbát ismertetjük.

#### **Kiugró érték vizsgálata Grubbs próbával**

A próba lehetőséget ad egy, vagy két legkisebb, illetve legnagyobb érték vizsgálatára.

Az n db vizsgálandó értéket növekvő sorrendbe állítva 1-től n-ig, attól függően, hogy egy vagy két kiugró értéket vizsgálunk, a következő számításmenetet alkalmazzuk.

#### Egy kiugró érték vizsgálata esetén

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 4.56$$
$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = 0.1205$$

Legnagyobb értéket vizsgálva:  $G_n = \frac{(x_n - \bar{x})}{s}$

Legkisebb értéket vizsgálva:  $G_1 = \frac{(\bar{x} - x_1)}{s}$

Amennyiben  $G$  számított értéke nagyobb, mint a Függelék 10. Táblázat egy kiugró értékre vonatkozó táblázatos értéke, a vizsgált adat az adott valószínűségi szinten kilóg a normális eloszlásból.

Esetünkben csupán a  $\lg N = 4.38$  érték gyanítható kiugróan kicsinek, a többi adathoz képest

Kiszámítva  $G_1$  értékét:

$$G_1 = \frac{(4.56 - 4.38)}{0.1205} = 1.440$$

$G_1$  számított értéke kisebb, mint az  $n=8$ ,  $\alpha=5\%$  paraméterekhez tartozó  $G_{kr} = 2.126$  érték, tehát a minták  $\lg N$  értékei homogénnek tekinthetők.

### Két kiugró érték vizsgálata esetén

Legnagyobb értékeket vizsgálva:

$$\bar{x}_{n-1,n} = \frac{1}{n-2} \sum_{i=1}^{n-2} x_i \quad s_0^2 = \sum_{i=1}^n (x_i - \bar{x})^2 \quad s_{n-1,n}^2 = \sum_{i=1}^{n-2} (x_i - \bar{x}_{n-1,n})^2$$

$$G = \frac{s_{n-1,n}^2}{s_0^2}$$

Legkisebb értékeket vizsgálva:

$$\bar{x}_{1,2} = \frac{1}{n-2} \sum_{i=3}^n x_i \quad s_{1,2}^2 = \sum_{i=3}^n (x_i - \bar{x}_{1,2})^2$$

$$G = \frac{s_{1,2}^2}{s_0^2}$$

Amennyiben a számított  $G$  érték kisebb, mint a Függelék 10. Táblázatban lévő kritikus érték, a két legnagyobb, vagy legkisebb adatot a normális eloszlásból kiugrónak tekintjük.

Tekintettel arra, hogy a mintaelemek egy kiugró adatot vizsgálva a fentiekben homogénnek bizonyultak, ezt a próbát nem végezzük el.

### ***Laboratóriumok közötti módszer összehasonlító vizsgálatok eredményeinek értékelése***

A vizsgálatra kiadott minták homogenitása alapján nem volt elvi akadálya a laboratóriumi eredmények kiértékelésének.

A laboratóriumok által mért sejtszámok logaritmikus transzformációja után elvégeztük az egyes laboratóriumok által megadott párhuzamos eredmények szórásának összehasonlítását.

#### **6./7. Táblázat.**

#### Laboratóriumi eredmények (lg N)

Laboratórium	lg N			Átlag	S <sup>2</sup>
1.	4,41	4,54	4,53	4,493	0,005233
2.	4,62	4,71	4,79	4,707	0,007233
3.	4,56	4,73	4,81	4,700	0,0163
4.	4,26	4,43	4,49	4,393	0,014233
5.	4,63	4,71	4,36	4,567	0,03363
6.	4,79	4,54	4,62	4,650	0,0163
7.	4,93	4,97	4,88	4,927	0,002033
8.	4,51	4,45	4,61	4,523	0,006533
9.	4,72	4,62	4,56	4,633	0,006533
10.	4,68	4,79	4,71	4,727	0,003233
11.	4,49	4,57	4,61	4,557	0,003733
12.	4,67	4,79	4,72	4,727	0,003633
13.	4,49	4,45	4,52	4,487	0,001233
14.	4,67	4,71	4,72	4,700	0,0007
15.	4,45	4,52	4,67	4,547	0,012633
16.	4,76	4,56	4,59	4,637	0,011633
17.	4,46	4,54	4,59	4,530	0,0043
18.	4,79	4,63	4,69	4,703	0,006533
19.	4,71	4,67	4,51	4,630	0,0112
20.	4,66	4,76	4,58	4,667	0,008133
<b>Összeg</b>				<b>92,505</b>	<b>0,174992</b>
<b>Átlag</b>				<b>4,6225</b>	<b>0,00875</b>
<b>Szórás</b>					<b>0.0935</b>

Megvizsgáltuk, hogy a laboratóriumokon belüli párhuzamos mérések szórásnégyzetei között nincs-e kiugróan nagy érték. Erre a célra az azonos számú párhuzamosok esetén használható Cochran-próbát alkalmaztuk.

#### **Kiugróan nagy szórásnégyzet vizsgálata Cochran próbával**

Próbastatisztika: 
$$C = \frac{s_{\max}^2}{\sum_{i=1}^p s_i^2},$$

Ahol  $s_{\max}^2$  a legnagyobb szórásnégyzet, p a laboratóriumok száma

Amennyiben C számított értéke kisebb, vagy egyenlő a Függelék 8. Táblázatban lévő kritikus értéknél, a szórások homogenitása az adott valószínűségi szinten elfogadható.

Kiszámítva C értékét:

$$C = \frac{0.0163}{0.174992} = \mathbf{0.0931}$$

A 20 laboratórium, 3 párhuzamos és  $\alpha=5\%$  elsőfajú hibához tartozó kritikus érték: 0.270, ennél a számított hányadosunk kisebb, tehát a szórások homogenitásának hipotézisét elfogadjuk.

Mivel a párhuzamos vizsgálatok szórásában kiugró értéket nem találtunk, a variancia-analízisben az összes laboratórium minden adatát felhasználtuk. (Kiugró adatok esetében az illető laboratóriumot ki kell hagyni az értékelésből.)

Az egyszempontos variancia-analízis számításmenete teljes mértékben megegyezik a 4.13. példában ismertetett eljárással. Az eredményeket a 6./8. Táblázatban foglaltuk össze.

**6./8. Táblázat.** Variancia-táblázat.

Variancia forrás	Négyzetössze g	Sz.. fok	S <sup>2</sup>	F	Szign. szint
Összes	1.1361	60-1=59			
Laboratóriumok között	0.7861	20-1=19	0.04137	4.73***	0.0000
Laboratóriumon belül	0.3500	40	0.00875 = S <sub>0</sub> <sup>2</sup>		

Az F-próba kritikus értékei a számláló 19, a nevező 40 szabadsági fokánál:  $F_{95\%}=1.85$ ,  $F_{99\%}=2.4$

A számított érték ennél jelentősen nagyobb, ezért megállapíthatjuk, hogy a laboratóriumok átlagértékei között szignifikáns különbség van.

A szignifikancia szint jelölésére a számított F érték után tett csillagok utalnak. Egy csillag 95, két csillag 99, három csillag 99.9%-os valószínűségű szignifikáns eltérésre utal.

Esetünkben a számítógépes program által megadott szignifikancia-szint 0.0000, ami azt jelenti, hogy a szignifikáns különbség valószínűsége 99.99% feletti (0.01% elsőfajú hiba megengedése már szignifikáns eltérést eredményez. Microsoft Excel programmal számolva a szignifikancia-határhoz tartozó elsőfajú hiba pontos értéke  $\alpha = 0.00177\%$ ).

Az átlagértékek közötti legkisebb szignifikáns különbség (LSD) számítása:

$$LSD = t \cdot \sqrt{S_0^2 \cdot \frac{2}{p}} = t \cdot S_0 \cdot \sqrt{\frac{2}{p}},$$

S<sub>0</sub><sup>2</sup> a véletlen szórásnégyzet, p pedig az összehasonlítandó kezeléseken belüli párhuzamosok száma. A táblázatos t érték szabadsági foka mindig a véletlen szórásnégyzet szabadsági foka.

A 40-es szabadsági fokhoz és 95% kétoldali valószínűségi szinthez tartozó érték:  $t = 2.02$ .

$$S_0 = \sqrt{0.00875} = 0.0935$$

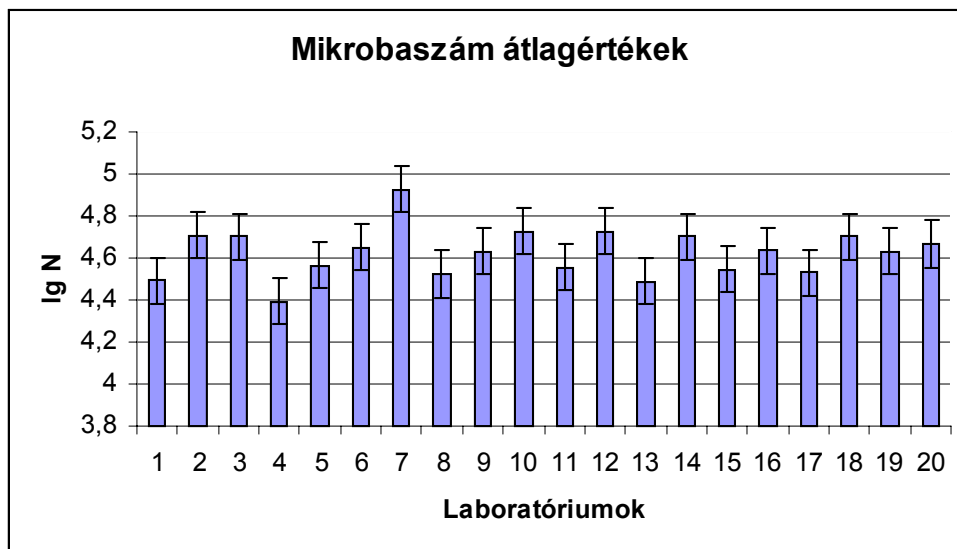
Az átlagértékek közötti legkisebb szignifikáns differencia:

$$\text{LSD} = 2.02 \cdot 0.0935 \cdot \sqrt{2/3} = \mathbf{0.1542}$$

Az átlagértékek 95%-os konfidencia-intervalluma:

$$\Delta = \pm t \cdot \frac{S_0}{\sqrt{n}} = \pm 2.02 \cdot \frac{0.0935}{\sqrt{3}} = \pm \mathbf{0.109}$$

A variancia-analízis eredményeit a 6./4. ábra szemlélteti, amelyen a különböző laboratóriumokhoz tartozó átlagértékeket és azok közös konfidencia-intervallumát tüntettük fel. A konfidencia-intervallumok esetenként átfedhetnek egymást, de ez még nem jelenti azt, hogy nincs szignifikáns különbség a két átlagérték között. A legkisebb szignifikáns különbség számított értéke az átlagértékek közötti különbségre vonatkozik és nem a konfidencia-határookra.



**6./4. ábra**

Mikrobaszám átlagértékek 95%-os konfidencia-intervallumai

A mikrobaszám meghatározás ismételhetőségét és reprodukálhatóságát a variancia-táblázatban összefoglalt értékekből határozhatjuk meg.

### Ismételhetőség

Az ismételhetőségre jellemző szórásnégyzet a variancia-analízisből számított maradék szórásnégyzet:

$$s_r^2 = S_0^2$$

$$S_r = \mathbf{0.0935}$$

Ismételhetőségi körülmények között, laboratóriumon belül megengedhető legnagyobb eltérés két mérés között ( $r$ ) a mérési eredmények legkisebb szignifikáns differenciája alapján számítható.

$$r = \Delta \lg N = t \cdot S_r \cdot \sqrt{2} = 2.02 \cdot 0.0935 \cdot \sqrt{2} = \mathbf{0.267}$$

### Reprodukálhatóság

A reprodukálhatóságra jellemző szórásnégyzet a laboratóriumi hatásra és a véletlen hatásra jellemző szórásnégyzetek összegeként határozható meg:

$$s_R^2 = s_L^2 + s_r^2$$

A laboratóriumok közötti szórásnégyzet variancia-analízisből meghatározott értékéből a laboratóriumi tiszta hatás a következő módon határozható meg:

$$s_L^2 = \frac{(s_{LK}^2 - s_r^2)}{p}$$

Ahol  $s_{LK}^2$  a variancia-táblázat Laboratóriumok közötti szórásnégyzete:  $s_{LK}^2 = 0.04137$

$s_r^2$  a variancia-analízis maradék szórásnégyzete:  $s_r^2 = 0.00875$

$p$  a laboratóriumokon belüli párhuzamosok száma:  $p = 3$

$$s_R^2 = s_L^2 + s_r^2 = \frac{(s_{LK}^2 - s_r^2)}{p} + s_r^2$$

Behelyettesítve az adatokat a variancia-táblázatból:

$$s_R^2 = \frac{(0.04137 - 0.00875)}{3} + 0.00875 = 0.0109 + 0.00875 = \mathbf{0.0196}$$

$$S_R = \mathbf{0.140}$$

Reprodukálhatósági körülmények között, két laboratórium mérése között megengedhető legnagyobb eltérés (**R**) a mérési eredmények legkisebb szignifikáns differenciája alapján számítható.

$$\mathbf{R} = \Delta \lg N = t \cdot S_R \cdot \sqrt{2} = 2.02 \cdot 0.140 \cdot \sqrt{2} = \mathbf{0.566}$$

Elméletileg előfordulhat olyan eset, hogy a laboratóriumok közötti szórás kisebb a véletlen szórásnál (a variancia-analízis F értéke kisebb 1-nél,  $s_L^2$  negatívnak adódna). Ilyenkor a véletlen hatások mellett a laboratóriumok által bevitt hiba elhanyagolható, az ismételhetőség és a reprodukálhatóság szórásnégyzetére egyaránt a maradék szórásnégyzet szolgál.





**MATEMATIKAI-STATISZTIKAI  
TÁBLÁZATOK**



**Függelék 1./a. Táblázat.** Standardizált normális eloszlás eloszlásfüggvénye.  $\Phi(-u) = 1 - \Phi(u)$ 

u	$\Phi(u)$	u	$\Phi(u)$	u	$\Phi(u)$	u	$\Phi(u)$	u	$\Phi(u)$
0.00	0.5000	0.30	0.6179	0.60	0.7257	0.90	0.8159	1.20	0.8840
0.01	0.5040	0.31	0.6217	0.61	0.7291	0.91	0.8186	1.21	0.8869
0.02	0.5080	0.32	0.6255	0.62	0.7324	0.92	0.8212	1.22	0.8888
0.03	0.5120	0.33	0.6293	0.63	0.7357	0.93	0.8238	1.23	0.8907
0.04	0.5160	0.34	0.6331	0.64	0.7389	0.94	0.8264	1.24	0.8925
0.05	0.5199	0.35	0.6368	0.65	0.7422	0.95	0.8289	1.25	0.8944
0.06	0.5239	0.36	0.6406	0.66	0.7454	0.96	0.8315	1.26	0.8962
0.07	0.5279	0.37	0.6443	0.67	0.7486	0.97	0.8340	1.27	0.8980
0.08	0.5319	0.38	0.6480	0.68	0.7517	0.98	0.8365	1.28	0.8997
0.09	0.5359	0.39	0.6517	0.69	0.7549	0.99	0.8389	1.29	0.9015
0.10	0.5398	0.40	0.6554	0.70	0.7580	1.00	0.8413	1.30	0.9032
0.11	0.5438	0.41	0.6591	0.71	0.7611	1.01	0.8438	1.31	0.9049
0.12	0.5478	0.42	0.6628	0.72	0.7642	1.02	0.8461	1.32	0.9066
0.13	0.5517	0.43	0.6664	0.73	0.7673	1.03	0.8485	1.33	0.9082
0.14	0.5557	0.44	0.6700	0.74	0.7703	1.04	0.8508	1.34	0.9099
0.15	0.5596	0.45	0.6736	0.75	0.7734	1.05	0.8531	1.35	0.9115
0.16	0.5636	0.46	0.6772	0.76	0.7764	1.06	0.8554	1.36	0.9131
0.17	0.5675	0.47	0.6808	0.77	0.7794	1.07	0.8577	1.37	0.9147
0.18	0.5714	0.48	0.6844	0.78	0.7823	1.08	0.8599	1.38	0.9162
0.19	0.5753	0.49	0.6879	0.79	0.7853	1.09	0.8621	1.39	0.9177
0.20	0.5793	0.50	0.6915	0.80	0.7881	1.10	0.8643	1.40	0.9192
0.21	0.5832	0.51	0.6950	0.81	0.7910	1.11	0.8655	1.41	0.9207
0.22	0.5871	0.52	0.6985	0.82	0.7939	1.12	0.8686	1.42	0.9222
0.23	0.5910	0.53	0.7019	0.83	0.7967	1.13	0.8708	1.43	0.9235
0.24	0.5948	0.54	0.7054	0.84	0.7995	1.14	0.8729	1.44	0.9251
0.25	0.5987	0.55	0.7088	0.85	0.8023	1.15	0.8749	1.45	0.9265
0.26	0.6026	0.56	0.7123	0.86	0.8051	1.16	0.8770	1.46	0.9279
0.27	0.6064	0.57	0.7157	0.87	0.8078	1.17	0.8790	1.47	0.9292
0.28	0.6103	0.58	0.7190	0.88	0.8106	1.18	0.8810	1.48	0.9306
0.29	0.6141	0.59	0.7224	0.89	0.8133	1.19	0.8830	1.49	0.9319

**Függelék 1./b. Táblázat.** Standardizált normális eloszlás eloszlásfüggvénye  $\Phi(-u) = 1 - \Phi(u)$ 

<b>u</b>	<b><math>\Phi(u)</math></b>	<b>u</b>	<b><math>\Phi(u)</math></b>	<b>u</b>	<b><math>\Phi(u)</math></b>	<b>u</b>	<b><math>\Phi(u)</math></b>	<b>u</b>	<b><math>\Phi(u)</math></b>
1.50	0.9332	1.80	0.9641	2.10	0.9821	2.40	0.9918	2.70	0.9965
1.51	0.9345	1.81	0.9649	2.11	0.9826	2.41	0.9920	2.71	0.9966
1.52	0.9357	1.82	0.9656	2.12	0.9830	2.42	0.9922	2.72	0.9967
1.53	0.9370	1.83	0.9664	2.13	0.9834	2.43	0.9925	2.73	0.9968
1.54	0.9382	1.84	0.9671	2.14	0.9838	2.44	0.9927	2.74	0.9969
1.55	0.9394	1.85	0.9678	2.15	0.9842	2.45	0.9929	2.75	0.9970
1.56	0.9406	1.86	0.9686	2.16	0.9846	2.46	0.9931	2.76	0.9971
1.57	0.9418	1.87	0.9693	2.17	0.9850	2.47	0.9932	2.77	0.9972
1.58	0.9429	1.88	0.9699	2.18	0.9854	2.48	0.9934	2.78	0.9973
1.59	0.9441	1.89	0.9806	2.19	0.9857	2.49	0.9936	2.79	0.9974
1.60	0.9452	1.90	0.9713	2.20	0.9861	2.50	0.9938	2.80	0.9974
1.61	0.9463	1.91	0.9719	2.21	0.9864	2.51	0.9940	2.81	0.9975
1.62	0.9474	1.92	0.9726	2.22	0.9868	2.52	0.9941	2.82	0.9976
1.63	0.9484	1.93	0.9732	2.23	0.9871	2.53	0.9943	2.83	0.9977
1.64	0.9495	1.94	0.9738	2.24	0.9875	2.54	0.9945	2.84	0.9977
1.65	0.9505	1.95	0.9744	2.25	0.9878	2.55	0.9946	2.85	0.9978
1.66	0.9515	1.96	0.9750	2.26	0.9881	2.56	0.9948	2.86	0.9979
1.67	0.9525	1.97	0.9756	2.27	0.9884	2.57	0.9949	2.87	0.9979
1.68	0.9535	1.98	0.9761	2.28	0.9887	2.58	0.9951	2.88	0.9980
1.69	0.9545	1.99	0.9767	2.29	0.9890	2.59	0.9952	2.89	0.9981
1.70	0.9554	2.00	0.9772	2.30	0.9893	2.60	0.9953	2.90	0.9981
1.71	0.9564	2.01	0.9778	2.31	0.9896	2.61	0.9955	2.92	0.9983
1.72	0.9572	2.02	0.9783	2.32	0.9898	2.62	0.9956	2.94	0.9984
1.73	0.9582	2.03	0.9788	2.33	0.9901	2.63	0.9957	2.96	0.9985
1.74	0.9591	2.04	0.9793	2.34	0.9904	2.64	0.9959	2.98	0.9986
1.75	0.9599	2.05	0.9798	2.35	0.9906	2.65	0.9960	3.00	0.9987
1.76	0.9608	2.06	0.9803	2.36	0.9909	2.66	0.9961	3.2	0.9993
1.77	0.9616	2.07	0.9808	2.37	0.9911	2.67	0.9962	3.4	0.9997
1.78	0.9625	2.08	0.9812	2.38	0.9913	2.68	0.9963	3.6	0.9998
1.79	0.9633	2.09	0.9817	2.39	0.9916	2.69	0.9964	3.8	0.9999

**Függelék 2. Táblázat.****Student féle t-eloszlás**

Szf	P Egyoldali kérdésfeltevés							
	75%	87.5%	95%	97.5%	99%	99.5%	99.9%	99.95%
1	1.000	2.41	6.31	12.7	31.82	63.7	318.3	637
2	0.816	1.60	2.92	4.30	6.97	9.92	22.33	31.6
3	0.765	1.42	2.35	3.18	4.54	5.84	10.22	12.9
4	0.741	1.34	2.13	2.78	3.75	4.60	7.17	8.61
5	0.727	1.30	2.01	2.57	3.37	4.03	5.89	6.86
6	0.718	1.27	1.94	2.45	3.14	3.71	5.21	5.96
7	0.711	1.25	1.89	2.36	3.00	3.50	4.79	5.40
8	0.706	1.24	1.86	2.31	2.90	3.36	4.50	5.04
9	0.703	1.23	1.83	2.26	2.82	3.25	4.30	4.78
10	0.700	1.22	1.81	2.23	2.76	3.17	4.14	4.59
11	0.697	1.21	1.80	2.20	2.72	3.11	4.03	4.44
12	0.695	1.21	1.78	2.18	2.68	3.05	3.93	4.32
13	0.694	1.20	1.77	2.16	2.65	3.01	3.85	4.22
14	0.692	1.20	1.76	2.14	2.62	2.98	3.79	4.14
15	0.691	1.20	1.75	2.13	2.60	2.95	3.73	4.07
16	0.690	1.19	1.75	2.21	2.58	2.92	3.69	4.01
17	0.689	1.19	1.74	2.11	2.57	2.90	3.65	3.96
18	0.688	1.19	1.73	2.10	2.55	2.88	3.61	3.92
19	0.688	1.19	1.73	2.09	2.54	2.86	3.58	3.88
20	0.687	1.18	1.73	2.09	2.53	2.85	3.55	3.85
21	0.686	1.18	1.72	2.08	2.52	2.83	3.53	3.82
22	0.686	1.18	1.72	2.07	2.51	2.82	3.51	3.79
23	0.685	1.18	1.71	2.06	2.50	2.81	3.49	3.77
24	0.685	1.18	1.71	2.06	2.49	2.80	3.47	3.74
25	0.684	1.18	1.71	2.06	2.48	2.79	3.45	3.72
26	0.684	1.18	1.71	2.06	2.48	2.78	3.44	3.71
27	0.684	1.18	1.71	2.05	2.47	2.77	3.42	3.69
28	0.683	1.17	1.70	2.05	2.47	2.76	3.41	3.67
29	0.683	1.17	1.70	2.05	2.46	2.76	3.40	3.66
30	0.683	1.17	1.70	2.04	2.46	2.75	3.39	3.65
40	0.681	1.17	1.68	2.02	2.42	2.70	3.31	3.55
60	0.679	1.16	1.67	2.00	2.39	2.66	3.23	3.46
120	0.677	1.16	1.66	1.98	2.36	2.62	3.17	3.37
	0.674	1.15	1.64	1.96	2.33	2.58	3.09	3.29
Szf	50%	75%	90%	95%	98%	99%	99,8%	99,9%
	P Kétoldali kérdésfeltevés							

**Függelék 3. Táblázat.** $\chi^2$  eloszlás $(P = 1 - \alpha)$ 

Szf.	P %								
	1	2.5	5	10	90	95	97.5	99	99.9
1	$1.5 \cdot 10^{-4}$	$9.8 \cdot 10^{-4}$	0.0039	0.0158	2.71	3.84	5.02	6.62	10.8
2	0.0201	0.051	0.103	0.211	4.61	5.99	7.38	9.21	13.8
3	0.115	0.216	0.352	0.854	6.25	7.81	9.35	11.3	16.3
4	0.297	0.484	0.711	1.06	7.78	9.49	11.1	13.3	18.5
5	0.554	0.831	1.15	1.61	9.24	11.1	12.8	15.1	20.5
6	0.872	1.24	1.64	2.20	10.6	12.6	14.4	16.8	22.5
7	1.24	1.69	2.17	2.83	12.0	14.1	16.0	18.3	24.3
8	1.65	2.18	2.73	3.49	13.4	15.5	17.5	20.1	26.1
9	2.09	2.70	3.33	4.17	14.7	16.9	19.0	21.7	27.9
10	2.56	3.25	3.94	4.87	16.0	18.3	20.5	23.2	29.6
11	3.05	3.82	4.57	5.58	17.3	19.7	21.9	24.7	31.3
12	3.57	4.40	5.23	6.30	18.5	21.0	23.3	26.2	32.9
13	4.11	5.01	5.89	7.04	19.8	22.4	24.7	27.7	34.5
14	4.66	5.63	6.57	7.79	21.1	23.7	26.1	29.1	36.1
15	5.23	6.26	7.26	8.55	22.3	25.0	27.5	30.6	37.7
16	5.81	6.91	7.96	9.31	23.5	26.3	28.8	32.0	39.3
17	6.41	7.56	8.67	10.1	24.6	27.6	30.2	33.4	40.8
18	7.01	8.23	9.39	10.9	26.0	28.9	31.5	34.8	42.3
19	7.63	8.91	10.1	11.7	27.2	30.1	32.9	36.2	43.8
20	8.26	9.50	10.9	12.4	28.4	31.4	34.2	37.6	45.3
21	8.90	10.3	11.6	13.2	29.6	32.7	35.5	38.9	46.8
22	9.54	11.0	12.3	14.0	30.8	33.9	36.8	40.3	48.3
23	10.2	11.7	13.1	14.8	32.0	35.2	38.1	41.6	49.7
24	10.9	12.4	13.8	15.7	33.2	36.4	39.4	43.0	51.0
25	11.5	13.1	14.6	16.5	34.4	37.7	40.6	44.3	52.6
26	12.2	13.8	15.4	17.3	35.6	38.9	41.9	45.6	53.1
27	12.9	14.6	16.2	18.1	36.7	40.1	43.2	47.0	55.5
28	13.6	15.3	16.9	18.9	37.9	41.3	44.5	48.3	56.9
29	14.3	16.0	17.7	19.8	39.1	42.6	45.7	49.6	58.3
30	15.0	16.8	18.5	20.6	40.3	43.8	47.0	50.9	59.7
40	22.2	24.4	26.5	29.1	51.8	55.8	59.3	63.7	73.4
50	29.7	32.4	34.8	37.7	63.2	67.5	71.4	76.2	86.7
60	37.5	40.5	43.2	46.5	74.4	79.1	83.3	88.4	99.6
70	45.4	48.8	51.7	55.3	85.5	90.5	95.0	100.4	112.3
80	53.5	57.2	60.4	64.3	96.6	101.9	106.6	112.3	124.8
90	61.8	65.6	69.1	73.3	107.6	113.1	118.1	124.1	137.2
100	70.1	74.2	77.9	82.4	118.5	124.3	129.6	135.8	149.4

**Függelék 4./a. Táblázat.****F táblázat. P = 95%**

Szf nevező	Számológó szabadsági foka												
	1	2	3	4	5	6	8	10	15	20	30	50	
<b>1</b>	161	200	216	225	230	234	239	242	246	248	250	252	254
<b>2</b>	19	19	19	19	19	19	19	19	19	19	20	20	20
<b>3</b>	10.1	9.6	9.3	9.1	9.0	8.9	8.8	8.8	8.7	8.7	8.6	8.6	8.5
<b>4</b>	7.7	6.9	6.6	4.4	6.3	6.2	6.0	6.0	5.9	5.8	5.8	5.7	5.6
<b>5</b>	6.6	5.8	5.4	5.2	5.1	5.0	4.8	4.7	4.6	4.6	4.5	4.4	4.4
<b>6</b>	6.0	5.1	4.8	4.5	4.4	4.3	4.2	4.1	3.9	3.9	3.8	3.8	3.7
<b>7</b>	5.6	4.7	4.4	4.1	4.0	3.9	3.7	3.6	3.5	3.4	3.4	3.3	3.2
<b>8</b>	5.3	4.5	4.1	3.8	3.7	3.6	3.4	3.4	3.2	3.2	3.1	3.0	2.9
<b>9</b>	5.1	4.3	3.9	3.6	3.5	3.4	3.2	3.1	3.0	2.9	2.9	2.8	2.7
<b>10</b>	5.0	4.1	3.7	3.5	3.3	3.2	3.1	3.0	2.9	2.8	2.7	2.6	2.5
<b>12</b>	4.8	3.9	3.5	3.3	3.1	3.0	2.9	2.8	2.6	2.5	2.5	2.4	2.3
<b>14</b>	4.6	3.7	3.3	3.1	3.0	2.9	2.7	2.6	2.5	2.4	2.3	2.2	2.1
<b>16</b>	4.5	3.6	3.2	3.0	2.9	2.7	2.6	2.5	2.4	2.3	2.2	2.1	2.0
<b>18</b>	4.4	3.6	3.2	2.9	2.8	2.7	2.5	2.4	2.3	2.2	2.1	2.0	1.9
<b>20</b>	4.4	3.5	3.1	2.9	2.7	2.6	2.5	2.4	2.2	2.1	2.0	2.0	1.8
<b>25</b>	4.2	3.4	3.0	2.8	2.6	2.5	2.3	2.2	2.1	2.0	1.9	1.8	1.7
<b>30</b>	4.2	3.3	2.9	2.7	2.5	2.4	2.3	2.2	2.0	1.9	1.8	1.8	1.6
<b>40</b>	4.1	3.3	2.8	2.6	2.5	2.3	2.2	2.1	1.9	1.8	1.7	1.7	1.5
<b>50</b>	4.0	3.2	2.8	2.6	2.4	2.3	2.1	2.0	1.9	1.8	1.7	1.6	1.4
<b>60</b>	4.0	3.2	2.8	2.5	2.4	2.3	2.1	2.0	1.8	1.8	1.7	1.6	1.4
<b>80</b>	4.0	3.1	2.7	2.5	2.3	2.2	2.1	2.0	1.8	1.7	1.6	1.5	1.3
<b>100</b>	3.9	3.1	2.7	2.5	2.3	2.2	2.0	1.9	1.8	1.7	1.6	1.5	1.3
	3.84	3.00	2.60	2.37	2.21	2.10	1.94	1.83	1.67	1.57	1.46	1.35	1.00

**Függelék 4./b. Táblázat.****F táblázat. P = 99%**

<b>Szf</b> <b>nevező</b>	<b>Számláló szabadsági foka</b>												
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>8</b>	<b>10</b>	<b>15</b>	<b>20</b>	<b>30</b>	<b>50</b>	
<b>1</b>	4100	5000	5400	5600	5800	5900	6000	6000	6200	6200	6200	6300	6400
<b>2</b>	98	99	99	99	99	99	99	99	99	99	99	100	100
<b>3</b>	34	31	29	29	28	28	27	27	27	27	27	26	26
<b>4</b>	21	18	17	16	16	15	15	15	14	14	14	13	13
<b>5</b>	16	13	12	11	11	11	10	10	9.6	9.6	9.4	9.2	9.0
<b>6</b>	14	11	9.8	8.8	8.8	8.5	8.1	7.9	7.6	7.4	7.2	7.1	6.9
<b>7</b>	12	9.6	8.5	7.8	7.5	7.2	6.8	6.6	9.3	6.2	6.0	5.9	5.7
<b>8</b>	11	8.7	7.6	7.0	6.6	6.4	6.0	5.8	5.5	5.4	5.2	5.1	4.9
<b>9</b>	11	8.0	7.0	6.4	6.1	5.8	5.5	5.3	5.0	4.8	4.7	5.5	4.3
<b>10</b>	10	7.6	6.6	6.0	5.6	5.4	5.1	4.9	4.7	4.4	4.3	4.1	3.9
<b>12</b>	9.3	6.9	6.0	5.4	5.1	4.8	4.5	4.3	4.0	3.9	3.7	3.6	3.4
<b>14</b>	8.9	6.5	5.6	5.0	4.7	4.5	4.1	3.9	3.7	3.5	3.4	3.2	3.0
<b>16</b>	8.5	6.2	5.3	4.8	4.4	4.2	3.9	3.7	3.4	3.3	3.1	3.0	2.8
<b>18</b>	8.3	6.0	5.1	4.6	4.3	4.0	3.7	3.5	3.2	3.1	2.9	2.8	2.6
<b>20</b>	8.1	5.9	4.9	4.4	4.1	3.9	3.6	3.4	3.1	2.9	2.8	2.6	2.4
<b>25</b>	7.8	5.6	4.7	4.2	3.9	3.6	3.3	3.1	2.9	2.7	2.5	2.4	2.2
<b>30</b>	7.6	5.4	4.5	4.0	3.7	3.5	3.2	3.0	2.7	2.6	2.4	2.3	2.0
<b>40</b>	7.3	5.2	4.3	3.8	3.5	3.3	3.0	2.8	2.5	2.4	2.2	2.1	1.8
<b>50</b>	7.2	5.1	4.2	3.7	3.4	3.2	2.9	2.7	2.4	2.3	2.1	2.0	1.7
<b>60</b>	7.1	5.0	4.1	3.7	3.3	3.1	2.8	2.6	2.4	2.2	2.0	1.9	1.6
<b>80</b>	7.0	4.9	4.0	3.6	3.3	3.0	2.7	2.6	2.3	2.1	1.9	1.8	1.5
<b>100</b>	6.09	4.8	4.0	3.5	3.2	3.0	2.7	2.5	2.2	2.1	1.9	1.7	1.4
	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.32	2.04	1.88	1.70	1.52	1.00



**Függelék 5./a. Táblázat. Binomiális eloszlás konfidencia-intervallum határai. P=95%**

k	n - k										
	0	1	2	3	4	5	6	7	8	9	10
<b>Alsó határ <math>p_1</math></b>											
0	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.025	0.013	0.008	0.006	0.005	0.004	0.004	0.003	0.003	0.003	0.002
2	0.158	0.094	0.068	0.053	0.043	0.037	0.032	0.028	0.025	0.023	0.021
3	0.292	0.194	0.147	0.118	0.099	0.085	0.075	0.067	0.060	0.055	0.050
4	0.398	0.284	0.223	0.184	0.157	0.137	0.122	0.109	0.099	0.091	0.084
5	0.478	0.359	0.290	0.245	0.212	0.187	0.167	0.151	0.139	0.128	0.118
6	0.541	0.421	0.349	0.299	0.262	0.234	0.211	0.192	0.177	0.163	0.152
7	0.590	0.473	0.400	0.348	0.308	0.277	0.251	0.230	0.213	0.198	0.184
8	0.631	0.517	0.444	0.390	0.349	0.316	0.289	0.266	0.247	0.230	0.215
9	0.664	0.555	0.482	0.428	0.386	0.351	0.323	0.299	0.278	0.260	0.244
10	0.692	0.587	0.516	0.462	0.419	0.384	0.354	0.329	0.308	0.289	0.272
11	0.715	0.615	0.545	0.492	0.449	0.413	0.383	0.357	0.335	0.315	0.298
12	0.735	0.640	0.572	0.519	0.476	0.440	0.410	0.384	0.361	0.340	0.322
13	0.753	0.661	0.595	0.544	0.501	0.465	0.435	0.408	0.384	0.364	0.345
14	0.768	0.681	0.617	0.566	0.524	0.488	0.457	0.430	0.407	0.385	0.366
15	0.782	0.698	0.636	0.586	0.544	0.509	0.478	0.451	0.427	0.406	0.386
16	0.794	0.713	0.653	0.604	0.563	0.529	0.498	0.471	0.447	0.425	0.405
17	0.805	0.727	0.669	0.621	0.581	0.547	0.516	0.488	0.465	0.443	0.423
18	0.815	0.740	0.683	0.637	0.597	0.564	0.533	0.506	0.482	0.460	0.440
19	0.824	0.751	0.696	0.651	0.612	0.579	0.549	0.522	0.508	0.467	0.456
20	0.832	0.762	0.708	0.664	0.626	0.593	0.564	0.537	0.513	0.492	0.472

k	n - k										
	0	1	2	3	4	5	6	7	8	9	10
<b>Felső határ <math>p_2</math></b>											
0	-	0.975	0.842	0.708	0.602	0.522	0.459	0.410	0.369	0.336	0.308
1	1.00	0.987	0.906	0.806	0.716	0.641	0.579	0.572	0.483	0.445	0.413
2	1.00	0.992	0.932	0.853	0.777	0.710	0.651	0.600	0.556	0.518	0.484
3	1.00	0.994	0.947	0.882	0.816	0.755	0.701	0.652	0.610	0.572	0.538
4	1.00	0.995	0.957	0.901	0.843	0.788	0.738	0.692	0.651	0.614	0.581
5	1.00	0.996	0.963	0.915	0.863	0.813	0.766	0.723	0.684	0.649	0.616
6	1.00	0.996	0.968	0.925	0.878	0.833	0.789	0.749	0.711	0.677	0.646
7	1.00	0.997	0.972	0.933	0.891	0.849	0.808	0.770	0.734	0.701	0.671
8	1.00	0.997	0.975	0.940	0.901	0.861	0.823	0.787	0.753	0.722	0.692
9	1.00	0.997	0.977	0.945	0.909	0.872	0.837	0.802	0.770	0.740	0.711
10	1.00	0.998	0.979	0.950	0.916	0.882	0.848	0.816	0.785	0.756	0.728
11	1.00	0.998	0.981	0.953	0.922	0.890	0.858	0.827	0.797	0.769	0.743
12	1.00	0.998	0.982	0.957	0.927	0.897	0.867	0.837	0.809	0.782	0.756
13	1.00	0.998	0.983	0.960	0.932	0.903	0.874	0.846	0.819	0.793	0.768
14	1.00	0.998	0.984	0.962	0.936	0.909	0.881	0.854	0.828	0.803	0.779
15	1.00	0.998	0.985	0.964	0.939	0.913	0.887	0.861	0.836	0.812	0.789
16	1.00	0.999	0.986	0.966	0.943	0.918	0.893	0.868	0.844	0.820	0.798
17	1.00	0.999	0.987	0.968	0.946	0.922	0.898	0.874	0.851	0.828	0.806
18	1.00	0.999	0.988	0.970	0.948	0.925	0.902	0.879	0.857	0.835	0.814
19	1.00	0.999	0.988	0.971	0.950	0.929	0.906	0.884	0.862	0.841	0.821
20	1.00	0.999	0.989	0.972	0.953	0.932	0.910	0.889	0.868	0.847	0.827

**Függelék 5./b. Táblázat.****Binomiális eloszlás konfidencia-intervallum határai. P=95%**

k	n - k									
	11	12	13	14	15	16	17	18	19	20
	<b>Alsó határ <math>p_1</math></b>									
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.001
2	0.019	0.018	0.017	0.016	0.015	0.014	0.013	0.012	0.012	0.011
3	0.047	0.043	0.040	0.038	0.036	0.034	0.032	0.030	0.029	0.028
4	0.078	0.073	0.068	0.064	0.061	0.057	0.054	0.052	0.050	0.047
5	0.110	0.103	0.097	0.091	0.087	0.082	0.078	0.075	0.071	0.068
6	0.142	0.133	0.126	0.119	0.113	0.107	0.102	0.098	0.094	0.090
7	0.173	0.163	0.154	0.146	0.139	0.132	0.126	0.121	0.116	0.111
8	0.203	0.191	0.181	0.172	0.164	0.156	0.149	0.143	0.138	0.132
9	0.231	0.218	0.207	0.197	0.188	0.180	0.172	0.165	0.159	0.153
10	0.257	0.244	0.232	0.221	0.211	0.202	0.194	0.186	0.179	0.173
11	0.282	0.268	0.266	0.244	0.234	0.224	0.215	0.207	0.199	0.192
12	0.326	0.291	0.278	0.266	0.255	0.245	0.235	0.227	0.218	0.211
13	0.328	0.313	0.299	0.287	0.275	0.264	0.255	0.245	0.237	0.229
14	0.349	0.334	0.320	0.306	0.295	0.283	0.273	0.264	0.255	0.247
15	0.369	0.353	0.339	0.325	0.313	0.302	0.291	0.281	0.271	0.263
16	0.388	0.372	0.357	0.343	0.331	0.319	0.308	0.298	0.288	0.280
17	0.406	0.389	0.374	0.360	0.347	0.335	0.324	0.314	0.304	0.295
18	0.422	0.406	0.391	0.376	0.363	0.351	0.340	0.329	0.319	0.310
19	0.439	0.422	0.408	0.392	0.379	0.366	0.355	0.344	0.334	0.324
20	0.454	0.437	0.421	0.407	0.393	0.381	0.369	0.358	0.348	0.338

k	n - k									
	11	12	13	14	15	16	17	18	19	20
	<b>Felső határ <math>p_2</math></b>									
0	0.285	0.265	0.247	0.232	0.218	0.206	0.195	0.185	0.176	0.168
1	0.385	0.360	0.339	0.319	0.302	0.287	0.273	0.260	0.249	0.238
2	0.445	0.428	0.405	0.383	0.364	0.347	0.331	0.317	0.304	0.292
3	0.508	0.481	0.456	0.434	0.414	0.396	0.379	0.363	0.349	0.336
4	0.551	0.524	0.499	0.476	0.456	0.437	0.419	0.403	0.388	0.374
5	0.587	0.560	0.535	0.512	0.491	0.471	0.453	0.436	0.421	0.407
6	0.617	0.590	0.565	0.543	0.522	0.502	0.484	0.467	0.451	0.436
7	0.643	0.616	0.592	0.670	0.549	0.529	0.512	0.494	0.478	0.463
8	0.665	0.639	0.616	0.593	0.573	0.553	0.535	0.518	0.502	0.487
9	0.685	0.660	0.636	0.615	0.594	0.575	0.557	0.540	0.524	0.508
10	0.702	0.678	0.655	0.634	0.614	0.595	0.577	0.560	0.544	0.528
11	0.718	0.694	0.672	0.651	0.631	0.612	0.594	0.578	0.561	0.546
12	0.732	0.709	0.687	0.666	0.647	0.628	0.611	0.594	0.578	0.563
13	0.744	0.722	0.701	0.680	0.661	0.643	0.626	0.609	0.594	0.579
14	0.756	0.734	0.713	0.694	0.675	0.657	0.640	0.624	0.608	0.593
15	0.766	0.745	0.725	0.705	0.687	0.669	0.653	0.637	0.621	0.607
16	0.776	0.755	0.736	0.717	0.698	0.681	0.665	0.649	0.634	0.619
17	0.785	0.765	0.745	0.727	0.709	0.692	0.676	0.660	0.645	0.631
18	0.793	0.773	0.755	0.736	0.719	0.702	0.686	0.671	0.656	0.642
19	0.801	0.782	0.763	0.745	0.728	0.712	0.696	0.681	0.666	0.652
20	0.808	0.789	0.771	0.753	0.737	0.720	0.705	0.690	0.676	0.662

**Függelék 6./a. Táblázat.****Poisson eloszlású gyakoriságok konfidencia határai. P=95%**

$k=\sum k_i$	Alsó határ	Felső határ	$k=\sum k_i$	Alsó határ	Felső határ	$k=\sum k_i$	Alsó határ	Felső határ
0	0.0	3.7	17	9.9	27.2	34	23.5	47.5
1	0.1	5.6	18	10.7	28.4	35	24.3	48.7
2	0.2	7.2	19	11.5	29.6	36	25.1	49.8
3	0.6	8.8	20	12.2	30.8	37	26.0	51.0
4	1.0	10.2	21	13.0	32.0	38	26.8	52.2
5	1.6	11.7	22	13.8	33.2	39	27.7	53.3
6	2.2	13.1	23	14.6	34.4	40	28.6	54.5
7	2.8	14.4	24	15.4	35.6	41	29.4	55.6
8	3.4	15.8	25	16.2	36.8	42	30.3	56.8
9	4.0	17.1	26	17.0	38.0	43	31.1	57.9
10	4.7	18.4	27	17.8	39.2	44	32.0	59.0
11	5.4	19.7	28	18.6	40.4	45	32.8	60.2
12	6.2	21.0	29	19.4	41.6	46	33.6	61.3
13	6.9	22.3	30	20.2	42.8	47	34.5	62.5
14	7.7	23.5	31	21.0	44.0	48	35.3	63.6
15	8.4	24.8	32	21.8	45.1	49	36.1	64.8
16	9.4	26.0	33	22.7	46.3	50	37.0	65.9

**Függelék 6./b. Táblázat. Poisson eloszlású gyakoriságok konfidencia határai. P=99%**

$k=\sum k_i$	Alsó határ	Felső határ	$k=\sum k_i$	Alsó határ	Felső határ	$k=\sum k_i$	Alsó határ	Felső határ
0	0.0	5.3	17	8.2	30.7	34	20.8	52.1
1	0.0	7.4	18	8.9	32.0	35	21.6	53.3
2	0.1	9.3	19	9.6	33.3	36	22.4	54.5
3	0.3	11.0	20	10.3	34.6	37	23.2	55.7
4	0.6	12.6	21	11.0	35.9	38	24.0	56.9
5	1.0	14.1	22	11.8	37.2	39	24.8	58.1
6	1.5	15.6	23	12.5	38.4	40	25.6	59.3
7	2.0	17.1	24	13.2	39.7	41	26.4	60.5
8	2.5	18.5	25	14.0	41.0	42	27.2	61.7
9	3.1	20.0	26	14.7	42.2	43	28.0	62.9
10	3.7	21.3	27	15.4	43.5	44	28.8	64.1
11	4.3	22.6	28	16.2	44.8	45	29.6	65.3
12	4.9	24.0	29	17.0	46.0	46	30.4	66.5
13	5.5	25.4	30	17.7	47.2	47	31.2	67.7
14	6.2	26.7	31	18.5	48.4	48	32.0	68.9
15	6.8	28.1	32	19.3	49.6	49	32.8	70.1
16	7.5	29.4	33	20.0	50.8	50	33.6	71.3

**Függelék 7. Táblázat.****A korrelációs együttható kritikus értékei**

Szabadsági fok	$\alpha$				
	0.1	0.05	0.02	0.01	0.001
1	0.98760	0.99692	0.999507	0.999877	0.9999988
2	0.9000	0.95000	0.98000	0.999000	0.999900
3	0.8054	0.8783	0.93433	0.95873	0.99116
4	0.7293	0.8114	0.8822	0.91720	0.97406
5	0.6694	0.7545	0.8329	0.8745	0.95074
6	0.6215	0.7067	0.7887	0.8343	0.92493
7	0.5822	0.6664	0.7498	0.7977	0.8982
8	0.5494	0.6319	0.7155	0.7646	0.8721
9	0.5214	0.6021	0.6851	0.7348	0.8471
10	0.4973	0.5760	0.6581	0.7079	0.8233
11	0.4762	0.5529	0.6339	0.6835	0.8010
12	0.4575	0.5324	0.6120	0.6614	0.7800
13	0.4409	0.5139	0.5923	0.6411	0.7603
14	0.4259	0.4973	0.5742	0.6226	0.7420
15	0.4124	0.4821	0.5577	0.6055	0.7246
16	0.4000	0.4683	0.5425	0.5897	0.7084
17	0.3887	0.4555	0.5285	0.5751	0.6932
18	0.3783	0.4438	0.5155	0.5614	0.6787
19	0.3687	0.4329	0.5034	0.5487	0.6652
20	0.3598	0.4227	0.4921	0.5368	0.6524
25	0.3233	0.3809	0.4451	0.4869	0.5874
30	0.2960	0.3494	0.4093	0.4487	0.5541
35	0.2746	0.3246	0.3810	0.4182	0.5189
40	0.2573	0.3044	0.3578	0.3932	0.4896
45	0.2428	0.2875	0.3384	0.3721	0.4648
50	0.2306	0.2732	0.3218	0.3541	0.4433
60	0.2108	0.2500	0.2948	0.3248	0.4078
70	0.1954	0.2319	0.2723	0.3017	0.3799
80	0.1829	0.2172	0.2565	0.2830	0.3568
90	0.1726	0.2050	0.2422	0.2673	0.3375
100	0.1638	0.1946	0.2301	0.2540	0.3211

**Függelék 8. Táblázat.**

**Cochran próba kritikus értékei**

p	n = 2		n = 3		n = 4		n = 5		n = 6	
	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
2	-	-	0.995	0.975	0.979	0.939	0.959	0.906	0.937	0.877
3	0.993	0.967	0.942	0.871	0.883	0.798	0.834	0.746	0.793	0.707
4	0.968	0.906	0.864	0.768	0.781	0.684	0.721	0.629	0.676	0.590
5	0.928	0.841	0.788	0.684	0.696	0.598	0.633	0.544	0.588	0.506
6	0.883	0.781	0.722	0.616	0.626	0.532	0.564	0.480	0.520	0.445
7	0.838	0.727	0.664	0.561	0.568	0.480	0.508	0.431	0.466	0.397
8	0.794	0.680	0.615	0.516	0.521	0.438	0.463	0.391	0.423	0.360
9	0.754	0.638	0.573	0.478	0.481	0.403	0.425	0.358	0.387	0.329
10	0.718	0.602	0.536	0.445	0.447	0.373	0.393	0.331	0.357	0.303
11	0.684	0.570	0.504	0.417	0.418	0.348	0.366	0.308	0.332	0.281
12	0.653	0.541	0.475	0.392	0.392	0.326	0.343	0.288	0.310	0.262
13	0.624	0.515	0.450	0.371	0.369	0.307	0.322	0.271	0.291	0.243
14	0.599	0.492	0.427	0.352	0.349	0.291	0.304	0.255	0.274	0.232
15	0.575	0.471	0.407	0.335	0.332	0.276	0.288	0.242	0.259	0.220
16	0.553	0.452	0.388	0.319	0.316	0.262	0.274	0.230	0.246	0.208
17	0.532	0.434	0.372	0.305	0.301	0.250	0.261	0.219	0.234	0.198
18	0.514	0.418	0.356	0.293	0.288	0.240	0.249	0.209	0.223	0.189
19	0.496	0.403	0.343	0.281	0.276	0.230	0.238	0.200	0.214	0.181
20	0.480	0.389	0.330	0.270	0.265	0.220	0.229	0.192	0.205	0.174
21	0.465	0.377	0.318	0.261	0.255	0.212	0.220	0.185	0.197	0.167
22	0.450	0.365	0.307	0.252	0.246	0.204	0.212	0.178	0.189	0.160
23	0.437	0.354	0.297	0.243	0.238	0.197	0.204	0.172	0.182	0.155
24	0.425	0.343	0.287	0.235	0.230	0.191	0.197	0.166	0.176	0.149
25	0.413	0.334	0.278	0.228	0.222	0.185	0.190	0.160	0.170	0.144
26	0.402	0.325	0.270	0.221	0.215	0.179	0.184	0.155	0.164	0.140
27	0.391	0.316	0.262	0.215	0.209	0.173	0.179	0.150	0.159	0.135
28	0.382	0.308	0.255	0.209	0.202	0.168	0.173	0.146	0.154	0.131
29	0.372	0.300	0.248	0.203	0.196	0.164	0.168	0.142	0.150	0.127
30	0.363	0.293	0.241	0.198	0.191	0.159	0.164	0.138	0.145	0.124
31	0.355	0.286	0.235	0.193	0.186	0.155	0.159	0.134	0.141	0.120
32	0.347	0.280	0.229	0.188	0.181	0.151	0.155	0.131	0.138	0.117
33	0.339	0.273	0.224	0.184	0.177	0.147	0.151	0.127	0.134	0.114
34	0.332	0.267	0.218	0.179	0.172	0.144	0.147	0.124	0.131	0.111
35	0.325	0.262	0.213	0.175	0.168	0.140	0.144	0.121	0.127	0.108
36	0.318	0.256	0.208	0.172	0.165	0.137	0.140	0.118	0.124	0.106
37	0.312	0.251	0.204	0.168	0.161	0.134	0.137	0.116	0.121	0.103
38	0.306	0.246	0.200	0.164	0.157	0.131	0.134	0.113	0.119	0.101
39	0.300	0.242	0.196	0.161	0.154	0.129	0.131	0.111	0.116	0.099
40	0.294	0.237	0.192	0.158	0.151	0.126	0.128	0.108	0.114	0.097

p = szórásnégyzetek száma.

n = egy szórásnégyzethez tartozó párhuzamosok száma.

**Függelék 9. Táblázat.**

**Dixon próba kritikus r értékei.**

Képletek	n	$\alpha$				
		10%	5%	2%	1%	0.5%
$r_{10} = \frac{X_1 - X_2}{X_1 - X_n}$	3	0.886	0.941	0.976	0.988	0.994
	4	0.679	0.765	0.846	0.889	0.926
	5	0.557	0.642	0.729	0.780	0.821
	6	0.482	0.560	0.644	0.698	0.740
	7	0.434	0.507	0.586	0.637	0.680
$r_{11} = \frac{X_1 - X_2}{X_1 - X_{n-1}}$	8	0.479	0.554	0.631	0.683	0.725
	9	0.441	0.512	0.587	0.635	0.677
	10	0.409	0.477	0.551	0.597	0.639
$r_{21} = \frac{X_1 - X_3}{X_1 - X_{n-1}}$	11	0.517	0.576	0.638	0.679	0.713
	12	0.490	0.546	0.605	0.642	0.675
	13	0.467	0.521	0.578	0.615	0.649
$r_{22} = \frac{X_1 - X_3}{X_1 - X_{n-2}}$	14	0.492	0.546	0.602	0.641	0.674
	15	0.472	0.525	0.579	0.616	0.647
	16	0.454	0.507	0.559	0.595	0.624
	17	0.438	0.490	0.542	0.577	0.605
	18	0.424	0.475	0.527	0.561	0.589
	19	0.412	0.462	0.514	0.547	0.575
	20	0.401	0.450	0.502	0.535	0.562
	21	0.391	0.440	0.491	0.524	0.551
	22	0.382	0.430	0.481	0.514	0.541
	23	0.374	0.421	0.472	0.505	0.532
	24	0.367	0.413	0.464	0.497	0.524
	25	0.360	0.406	0.457	0.489	0.516

Az n db adat sorba rendezve  $X_1$ -től  $X_n$ -ig.

$X_1$  a kiugró (legnagyobb, vagy legkisebb) érték.

Amennyiben a számított r értéke nagyobb a táblázatos értéknél,  $X_1$  a feltételezett normális eloszlásból kiugró adatnak tekinthető.

**Függelék 10. Táblázat.****Grubbs próba kritikus értékei.**

n	Egy kiugró érték		Két kiugró érték	
	$\alpha = 1\%$ felső	$\alpha = 5\%$ felső	$\alpha = 1\%$ alsó	$\alpha = 5\%$ alsó
3	1.155	1.155	-	-
4	1.496	1.481	0.0000	0.0002
5	1.764	1.715	0.0018	0.0090
6	1.973	1.887	0.0116	0.0349
7	2.139	2.020	0.0308	0.0708
8	2.274	2.126	0.0563	0.1101
9	2.387	2.215	0.0851	0.1492
10	2.482	2.290	0.1150	0.1864
11	2.564	2.355	0.1448	0.2213
12	2.636	2.412	0.1738	0.2537
13	2.699	2.462	0.2016	0.2836
14	2.755	2.507	0.2280	0.3112
15	2.806	2.549	0.2530	0.3367
16	2.852	2.585	0.2767	0.3603
17	2.894	2.620	0.2990	0.3822
18	2.932	2.651	0.3200	0.4025
19	2.968	2.681	0.3398	0.4214
20	3.001	2.709	0.3585	0.4391
21	3.031	2.733	0.3761	0.4556
22	3.060	2.758	0.3927	0.4711
23	3.087	2.781	0.4085	0.4857
24	3.112	2.802	0.4234	0.4994
25	3.135	2.822	0.4376	0.5123
26	3.157	2.841	0.4510	0.5245
27	3.178	2.859	0.4638	0.5360
28	3.199	2.876	0.4759	0.5470
29	3.218	2.893	0.4875	0.5574
30	3.236	2.908	0.4985	0.5672
31	3.253	2.924	0.5091	0.5766
32	3.270	2.938	0.5192	0.5856
33	3.286	2.952	0.5288	0.5941
34	3.301	2.965	0.5381	0.6023
35	3.316	2.979	0.5469	0.6101
36	3.330	2.991	0.5554	0.6175
37	3.343	3.003	0.5636	0.6247
38	3.356	3.014	0.5714	0.6316
39	3.369	3.025	0.5789	0.6382
40	3.381	3.036	0.5862	0.6445

Egy kiugró érték esetén:

kiugró, ha a számított érték nagyobb, mint a táblázatos érték.

Két kiugró érték esetén:

kiugró, ha a számított érték kisebb, mint a táblázatos érték.





## FELHASZNÁLT IRODALOM

- Bányainé dr. Sándor Julianna, Perczelné dr. Zalai Magdolna (1983):  
Tartósított termékek statisztikai minőségellenőrzése.  
Mezőgazdasági Kiadó, Budapest.
- Deák, T., Lendvai, I., Reichart O. (1980): Mikrobiológiai gyártásellenőrzés és minőség.  
in: Deák, T., Farkas, J., Incze K.: Konzerv-, hús- és hűtőipari mikrobiológia.  
Mezőgazdasági Kiadó, Budapest.
- Reichart O. (1991): Some remarks on the bias of the MPN method  
*International Journal of Food Microbiology*. 13. 131-142.
- Sváb J.(1973): Biometriai módszerek a kutatásban.  
Mezőgazdasági Kiadó, Budapest.
- Zukál, E., Fényes, T., Körmendy L. (1969): A matematikai statisztika alapjai.  
in: Deák, T., Novák E. szerk.: Kísérletek tervezése és értékelése.  
Magyar Kémikusok Egyesülete.
- Lukács, O. (1987): Matematikai statisztika. Példatár.  
Műszaki Könyvkiadó, Budapest.
- Balogh, A., Dukáti, F., Sallay L. (1980): Minőségellenőrzés és megbízhatóság.  
Műszaki Könyvkiadó, Budapest.
- Korn, G. A., Korn, T. M. (1975): Matematikai Kézikönyv műszakiaknak.  
Műszaki Könyvkiadó, Budapest.
- ISO 5725-2: 1994. Accuracy (trueness and precision) of measurement methods and results -  
Part 2: Basic method for the determination of repeatability and  
reproducibility of a standard measurement method.