

## TÉMAVÁZLAT 4-7. ALKALOM

*Kémiai Számítástechnika Gyakorlat, Kémia BSc I. évf. 2019/2020 I. félév*

(összeállította: Tóth Gergely)

### STATISZTIKAI ALAPOK

Célja: egy halmazból, sokaságból kiválasztott minta alapján az egész halmazra vonatkozó következtetéseket vonjunk le. Az események eloszlását a véges minta miatt nem ismerjük tökéletesen, miként alkalmazzuk a valószínűségszámítás fogalmait ilyen esetben.

#### Várható érték becslése

várható érték  $N$  elemű mintára ( $y_i$  elemek):  $E(\xi) \approx \bar{y} = \frac{\sum_{i=1}^N y_i}{N}$  (ÁTLAG)

medián ( $y_{\text{medián}}$ ): középső érték, vagy két középső átlaga. Kevésbé érzékeny a kilógó (elszúrt) adatra.

MEDIÁN

módusz: leggyakoribb adat (diszkrét eloszlásnál értelmes) MÓDUSZ

további EXCEL függvények: MIN, MAX, KICSI, NAGY

#### Szórás és szórásnégyzet (variancia) becslése mintából

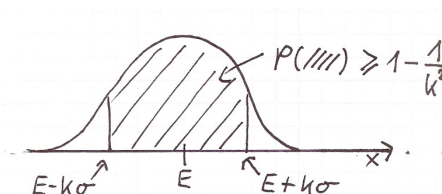
$$\sigma^2 \approx s^2 = \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N-1} \quad \sigma \approx s = \sqrt{\frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N-1}}$$

neve: becült szórás(négyzet), korrigált tapasztalati szórás(négyzet), VAR, SZÓRÁS

$\sigma^2$  becslése csak  $N$ -es osztás esetén „torzított” lenne, ( $N-1$ ) osztás esetén torzítatlan (Bessel-féle korrekció). Sajnos  $\sigma$  becslése így is torzított, tehát az igazi statisztikus varianciákkal és nem szórásokkal dolgozik!

#### Kilógó mérési adat kiválasztása

Háttér a Csebisev-egyenlőtlenség,



Csebisev-egyenlőtlenség szemléltetése tetszőleges eloszlásra

Szimmetrikus intervallumba esési valószínűségek a  
Csebisev-egyenlőtlenség alapján, illetve a normális eloszlásra

intervallum	tetszőleges eloszlásra P	normális eloszlásra P
$E \pm 1\sigma$		$P=0,682$
$E \pm 2\sigma$	$0,75 \leq P$	$P=0,954$
$E \pm 3\sigma$	$0,88 \leq P$	$P=0,997$

Normális eloszlásnál  $P=0,95$  esetén a szorzó 1,96.

Ha a mintában lehet hiba (pl. félremért kilógó adat), az átlagot célszerű a mediánnal becsülni, a terjedelmet az ún. kvartilisok segítségével.

alsó kvartilis ( $y_{1/4}$ ): az elem, aminél az adatok negyede kisebb, háromnegyede nagyobb

felső kvartilis ( $y_{3/4}$ ): az elem, aminél az adatok háromnegyede kisebb, negyede nagyobb

interkvartilis távolság:  $y_{3/4} - y_{1/4}$

gyanúsak (outlier) vagy eldobhatóak azok a kilógó adatok, amik kívül vannak a

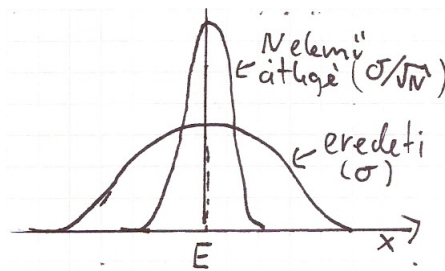
$y_{\text{medián}} \pm 1,5 * (y_{3/4} - y_{1/4})$ , esetleg a  $y_{\text{medián}} \pm 2,25 * (y_{3/4} - y_{1/4})$  intervallumon

**KVARTILIS, PERCENTILIS**

### Várható érték szórása

Vegyünk  $N$  elemű mintát egy  $E$  várható értékű és  $\sigma$  szórású eloszlásból, számoljuk ki  $\bar{y}$ -t. Ismételjük ezt meg sokszor. Mi lesz a számolt  $\bar{y}$ -k szórása ( $\sigma_N$ -nel jelölve)?

Bemutatható, hogy  $\sigma_N = \sigma / \sqrt{N}$ . Ugyanez érvényes a becült szórásokra is. Tehát:  $\lim_{N \rightarrow \infty} \sigma_N = 0$



Egy eredeti sokaság és az abból képzett  $N$  elemű átlagok sűrűségfüggvényei

### Várható érték megbízhatósági intervalluma

$N$  mérés  $\rightarrow \bar{y}$  Mit írunk le?  $\bar{y} \pm$  valamit, úgy, hogy tükrözze a várható érték pontosságát!

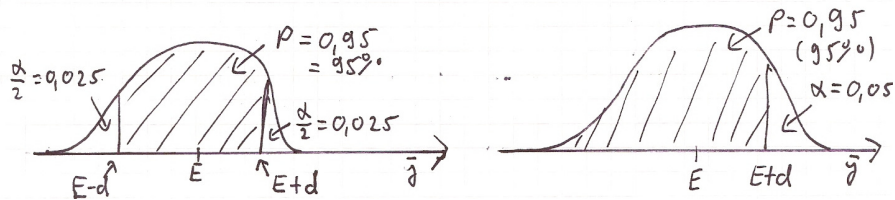
Ugyanaz az átlaga a két mérési sornak, de ugyanazt íránk le?

a) 10,001; 10,002; 10,000; 9,999; 9,998

b) 10,000; 7,000; 13,000; 9,000; 11,000

Megbízhatósági (konfidencia) intervallumokat adjunk meg a várható értékre:

Olyant, ahol  $P((\bar{y} - d) < E < (\bar{y} + d))$  valószínűsége nagyobb legyen, mint mondjuk 90%, vagy 95%, vagy 99%. Többnyire kétoldali intervallumot adunk meg, de lehet csak egyoldali is! Sokszor nem a minimális P-t hanem  $\alpha = 1 - P$  szignifikancia szintet adják meg.

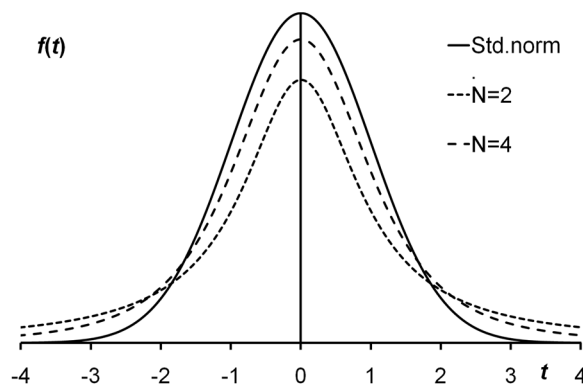


kétoldali és egyoldali megbízhatósági intervallumok szemléltetése

A ma elfogadott megoldás (Gosset="Student" 1908, Fisher 1925): t-eloszlás

$$t = \frac{\bar{y} - E}{s / \sqrt{N}}, \text{ ahol } \bar{y} \text{ és } s \text{ az } N \text{ elemű normális eloszlású mintából számolt várható érték és becslt}$$

szórás,  $E$  a sokaság (elméleti) várható értéke.  $t$  eloszlása kis  $N$ -re nem standard normális eloszlást ad, hanem ún.  $(N-1)$  szabadsági fokú  $t$ -eloszlást (más néven student-eloszlást). Szabadsági fok  $\approx$  független adatok száma. Ha megjelenik egy az adatokat összekötő egyenlet (pl. várható érték számolása miatt), az csökkenti a szabadsági fokok számát.



$t$ -eloszlás ( $N=2$ -re és  $N=4$ -re) és a standard normális eloszlás

Tehát várható érték megadása konfidencia intervallumával együtt:

$$\text{kétoldali konfidencia intervallummal: } \bar{y} \pm t_{\text{inverz}}(\alpha/2; N-1) \frac{s}{\sqrt{N}}, \text{ ahol } t_{\text{inverz}}(\alpha/2; N-1) \text{ azt az értéket}$$

szolgáltatja, hogy a valószínűségi változó milyen értékénél lesz az  $N-1$  szabadsági fokú  $t$ -eloszlás eloszlás függvényének értéke  $1 - \alpha/2$

$$\text{egyoldali konfidencia intervallumnál pl. csak a felső érték: } \bar{y} + t_{\text{inverz}}(\alpha; N-1) \frac{s}{\sqrt{N}}$$

$\pm$  rész számolása kétoldalinal: MEGBÍZHATÓSÁG.T függvényvel ( $\alpha$ ; szórás; méret)

+ vagy - rész számolása egyoldalinal: MEGBÍZHATÓSÁG.T függvényvel ( $\alpha$  kétszerese 2-vel; szórás; méret)

$\pm t$  nem értelmezi az EXCEL, tehát külön-külön cellába kerüljön  $\bar{y}$  és a  $\pm$  utáni rész!

$30 < N$  esetén a  $t$ -eloszlás közelíthető a standard normális eloszlással (MEGBÍZHATÓSÁG, MEGBÍZHATÓSÁG.NORM), tehát kétoldali 95%-os intervallumnál  $t_{\text{inverz}}(0,025;N-1)$  helyett számolhatunk pl. 1,96-tal.

*Feladatok:*

*Egy termék esetén a következő tömegeket mértük g-ban: 7,7; 14,0; 4,2; 7,8; 4,7; 9,1; 9,9; 7,9; 13,2; 4,6; 6,5; 6,6. Számoljuk ki a 95%-os kétoldali konfidencia intervallumot az átlagos tömegre!*

*A következő koncentrációkat mértük mol/dm<sup>3</sup> egységben: 0,120; 0,130; 0,110; 0,105; 0,125; 0,162; 0,135. Számoljuk ki a 95%-os kétoldali konfidencia intervallumot az átlagos tömegre! Határozzuk meg a minimális illetve a maximális értékét az átlagos tömegnek 95%-os egyoldali konfidencia sávval!*

## NEMLINEÁRIS EGYENLET MEGOLDÁSA, MAXIMUM ÉS MINIMUM KERESÉSE

Numerikus módszerek és eljárások szerepe

Iteratív módszerek

kezdőérték(ek) megadása

konvergencia, divergencia

leállási kritériumok:  $|x_{k+1}-x_k| < \epsilon$ , vagy  $|f(x_{k+1})-f(x_k)| < \epsilon$ ,

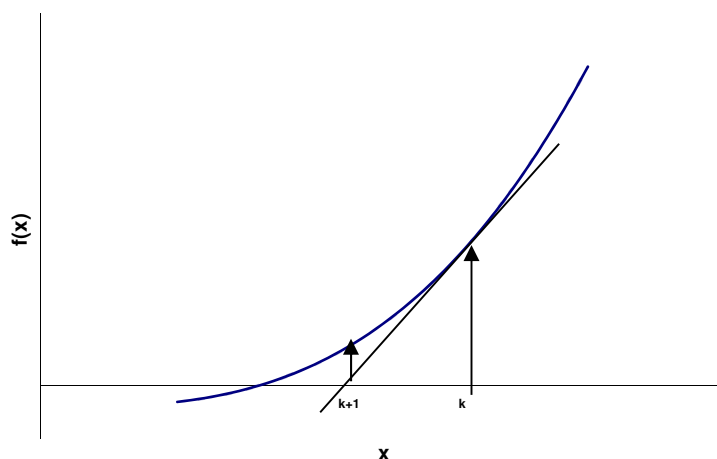
### Nemlineáris egyenlet numerikus megoldása

Feladat:  $y=h(x)$  függvény esetén egy adott  $y_0$  értékhez  $x_0$  meghatározása, ha  $x$  nem fejezhető ki explicit módon, mint  $x=g(y)$ .

Átrendezés  $f(x)=h(x)-y_0=0$  (Excellel való megoldásnál az átrendezés nem szükséges)

A sok közül egy alapszámítás:

#### Newton módszer



1 kezdőpont, divergálhat, iterációs

lépés:

$$x_{k+1} = x_k - f(x_k) / f'(x_k)$$

Egy másik alapszámítás: intervallum felezése (2 kezdőpont, biztos megoldás)

$x_a$  és  $x_f$  pontok választása úgy, hogy  $f(x_a) \cdot f(x_f) < 0$

$x_k = (x_a + x_f) / 2$ , ha  $f(x_k) \cdot f(x_a) < 0$ , új  $x_f = x_k$ , ellenkező esetben  $x_a = x_k$ ; Addig ismételjük, amíg  $[x_a, x_f] < \varepsilon$

Megoldás EXCEL-lel:

Adatok / Solver (esetleg aktiválni kell)

változó cella ( $x$  adott kezdőértékkel), célértékcella ( $f(x)$ ) kitöltése ( $x$ -nél annak a cellájára mutasson)

Minimum/maximum/értékkeresés beállítása

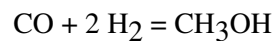
Korlátozó feltételek

*Feladatok:*

*Lennard-Jones párkölcsönhatási potenciál minimumának és  $x$ -tengellyel való metszéspontjának a meghatározása.*

### Kémiai egyensúly számítása

A metanol szintézise 25 % CO, 55 % H<sub>2</sub> és 20 % inert gáz összetételű elegyből indul ki (az adatok mol %-ban értendők). A



reakció parciális nyomásokkal kifejezett egyensúlyi állandója 350 °C hőmérsékleten:

$$K_p = \frac{P_{\text{CH}_3\text{OH}}}{P_{\text{CO}} P_{\text{H}_2}^2} = 1.4 \cdot 10^{-14} \text{ Pa}^{-2} \quad (1)$$

**Feladat:** Határozza meg az egyensúlyi összetételt!

Legyen  $x$  az egyensúlyi konverzió, és 100 mol elegyből induljunk ki. Írjuk fel a komponensek és az elegy kiindulási, ill. egyensúlyi anyagmennyiségeit:

Komponens	Kezdetben (mol)	Egyensúlyban (mol)
CO	25	25-25x
H <sub>2</sub>	55	55-50x
CH <sub>3</sub> OH	0	25x
inert	20	20
összesen	100	100-50x

A parciális nyomások az egyensúlyban:

$$P_{\text{CH}_3\text{OH}} = \frac{25x}{100 - 50x} P$$

$$P_{\text{CO}} = \frac{25 - 25x}{100 - 50x} P$$

$$P_{\text{H}_2} = \frac{55 - 50x}{100 - 50x} P$$

ahol  $P$  az össznyomás. Határozza meg az EXCEL segítségével az  $x$  egyensúlyi konverziót az (1) egyenlet alapján!

A következő össznyomás értékekkel számoljon:

- 1)  $P = 3.0 \cdot 10^7 \text{ Pa}$
- 2)  $P = 3.5 \cdot 10^7 \text{ Pa}$
- 3)  $P = 2.0 \cdot 10^7 \text{ Pa}$
- 4)  $P = 2.5 \cdot 10^7 \text{ Pa}$

A  $10^{-14}$ -en nagyságrendtől szabaduljon meg a nyomás  $10^7$ -jeivel való egyszerűsítésével!

#### Disszociációfok meghatározása

A nitrogén-dioxid disszociációja a  $2 \text{ NO}_2 \rightleftharpoons 2 \text{ NO} + \text{ O}_2$  egyenlet szerint megy végbe.

Ha a reakció *állandó térfogaton* játszódik le, akkor a komponensek egyensúlyi parciális nyomását fejezzük ki a nitrogén-dioxid kezdeti nyomásával ( $p_{\text{NO}_2}^0$ ) és disszociációfokával ( $\alpha$ ):

$$p_{\text{NO}_2} = (1 - \alpha) p_{\text{NO}_2}^0$$

$$p_{\text{NO}} = \alpha p_{\text{NO}_2}^0$$

$$p_{\text{O}_2} = 0.5\alpha p_{\text{NO}_2}^0$$

Helyettesítsük be ezeket a kifejezéseket az egyensúlyi állandó parciális nyomásokkal kifejezett alakjába:

$$K_a = K_p \left[ \frac{1}{p^\ominus} \right]^{\Delta v} = \frac{0.5\alpha p_{\text{NO}_2}^0 (\alpha p_{\text{NO}_2}^0)^2}{(1 - \alpha)^2 (p_{\text{NO}_2}^0)^2} \frac{1}{p^\ominus} \quad (1)$$

ahol  $p^\ominus = 100 \text{ kPa}$  a standard nyomás,  $\Delta v$  pedig a reakcióval járó sztöchiometriai szám változás, esetünkben 1. Az (1) egyenlet  $\alpha$ -ra megoldható.

Számítsa ki a disszociációfokot!

150 kPa nyomású tiszta nitrogén-dioxidból induljon ki, a hőmérséklet 700 K. Az egyensúlyi állandó ezen a hőmérsékleten:  $K_a = 0,18$

## STATISZTIKAI PRÓBÁK

Adott néhány adat(sor), ezek alapján kell eldönteni egy állításról, hogy igaz-e. Adatsor: valószínűségi változóra vonatkozik → a döntés valamilyen valószínűséggel lesz helyes. Állítás = hipotézis ( $H_0$ ), tagadása = ellenhipotézis ( $H_a$ ): a kettő közül kell döntenünk. Általános módszer: A feltett állítás általánosan leírható egy eloszlással. Azt nézzük, hová esik az adott adat(sor)-ra számítható érték.

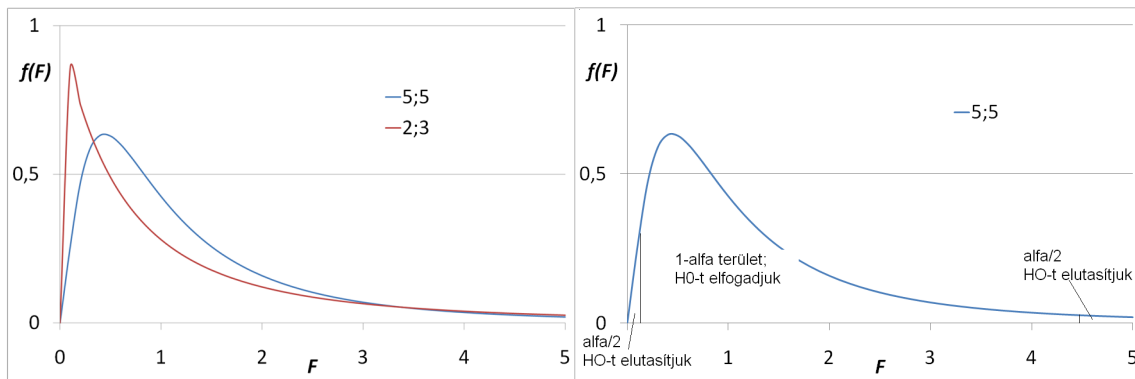
### F-próba – tekinthető a két minta alapján két szórás azonosnak

$H_0: \sigma_1 = \sigma_2$

$H_a: \sigma_1 \neq \sigma_2$ , ahol az elsőre  $N_1$  a másodikra  $N_2$  mérést végeztünk.

Pl. két műszer, két hallgató, két módszer: egyformán pontosak-e?

Háttér: F-eloszlás írja le két szórásnégyzet hányadosát, ha mindkét minta külön-külön normális eloszlásból származik. Két szabadsági fok:  $N_1-1$  és  $N_2-1$ . Becsült szórások a mintából:  $s_1$  és  $s_2$ .



Az F-eloszlás sűrűségfüggvénye két szabadságfok-párra

kétoldali intervallum alapján való hipotézisvizsgálat

Elvi megoldás:  $\xi_F = \frac{s_1^2}{s_2^2}$  kiszámolása, majd az  $N_1-1, N_2-1$  szabadsági fokú F-eloszlás alapján

megnézni, hová esik a számított érték. Feltételezünk adott  $\alpha$ -t (pl. 0.05).

F.PRÓB vagy F.PRÓBA (tömb1, tömb2) közvetlenül  $H_0$  „valószínűségét” számolja:

ha  $\alpha \leq F.PRÓBA$ ,  $H_0$ -t elfogadjuk

ha  $F.PRÓBA < \alpha$ ,  $H_0$ -t elutasítjuk ( $H_a$ -t fogadjuk el).

### Egymintás t-próba – a minta sokaságának várható értéke és egy elméleti várhatóérték tekinthető-e azonosnak

$E$  - minta sokaságának várható értéke

$\hat{y}$  - mintából számolt átlag = becsült várható érték

$E_0$  – elméleti várható érték (pl. hivatalos adat...)

$H_0: E=E_0$        $H_a: E \neq E_0$

Háttér:  $\xi = \frac{\hat{y} - E_0}{s / \sqrt{N}}$        $N-1$  szabadsági fokú  $t$ -eloszlást követ

Elvi megoldás:  $\xi = \frac{\hat{y} - E_0}{s/\sqrt{N}}$  kiszámolása és  $F(\xi)$  összevetése  $\alpha$ -val.

Kétoldali: ha  $\alpha/2 \leq F(\xi) \leq (1-\alpha/2)$ ,  $H_0$ -t elfogadjuk

ha  $F(\xi) < \alpha/2$  vagy  $(1-\alpha/2) < F(\xi)$ ,  $H_0$ -t elutasítjuk ( $H_a$ -t fogadjuk el)

Egyoldali, pl. csak a felfele kilógás „rossz”: ha  $F(\xi) \leq (1-\alpha)$ ,  $H_0$ -t elfogadjuk

ha  $(1-\alpha) < F(\xi)$ ,  $H_0$ -t elutasítjuk ( $H_a$ -t fogadjuk el)

Megoldás EXCEL-lel kétoldalira:

ha  $\alpha/2 \leq T.ELOSZL(\xi; N-1; 1) \leq (1-\alpha/2)$ ,  $H_0$ -t elfogadjuk

ha kívül esik,  $H_0$ -t elutasítjuk ( $H_a$ -t fogadjuk el)

$30 < N$ , lehet normál eloszlással dolgozni t-eloszlás helyett:

ha  $\alpha/2 \leq Z.PRÓBA(adattömb, E_0) \leq (1-\alpha/2)$ ,  $H_0$ -t elfogadjuk

ha kívül esik,  $H_0$ -t elutasítjuk ( $H_a$ -t fogadjuk el)

### Kétmintás t-próba – két minta sokaságának várható értéke tekinthető-e azonosnak

$E_1, E_2$  - két sokaság várható értéke

$\hat{y}_1, \hat{y}_2$  - mintából számolt átlagok = becsült várható értékek

**$H_0: E_1=E_2$        $H_a: E_1 \neq E_2$**

Általános képlet:  $\xi = (\text{becsült paraméter} - \text{elméleti paraméter}) / (\text{becsült szórása a paraméternek})$

becsült paraméter:  $\hat{y}_1 - \hat{y}_2$ ,      elméleti paraméter:  $E_1 - E_2$

szórás:  $s = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$ , ha  $30 < N_1, N_2$

helyette esetleg „pooled” variancia, ha  $N_1, N_2 \leq 30$

$$s_p^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \quad s = s_p \left( \frac{1}{\sqrt{N_1}} + \frac{1}{\sqrt{N_2}} \right)$$

Számolás, elfogadás/elutasítás ahogy az egymintás t-próbánál, vagy

ha  $\alpha \leq T.PRÓBA$  vagy  $T.PRÓB(adattömb_1; adattömb_2; 2; 2$  vagy  $3)$ ,  $H_0$ -t elfogadjuk

ha  $T.PRÓBA(adattömb; adattömb_2; 2; 2$  vagy  $3) < \alpha$ ,  $H_0$ -t elutasítjuk ( $H_a$ -t fogadjuk el)

### $\chi^2$ -próba – illeszkedés vizsgálata

$\chi^2$ -eloszlás:  $\xi_1, \xi_2, \xi_3, \xi_4, \dots$  standard normális eloszlású, akkor  $\xi = \xi_1^2 + \xi_2^2 + \dots + \xi_N^2$   $N$  szabadsági

fokú  $\chi^2$ -eloszlást követ. Ha több adat van és centrálunk, akkor a szabadsági fok =  $N-1$ .

$E(\xi) = \text{szabadsági fokok száma}$

$\sigma^2(\xi) = 2 * \text{szabadsági fokok száma}$

Kapcsolat részecskék energiájának eloszlásával: részecske  $v_x, v_y, v_z$  sebességei normál eloszlásúak

$v^2 = v_x^2 + v_y^2 + v_z^2 \rightarrow E_{\text{kinetikus}} = mv^2/2 \rightarrow \text{Maxwell-Boltzmann eloszlást követ} = \chi^2\text{-eloszlás } N=3$



$\chi^2$ -próba mire jó? Megnézni, hogy két görbe közötti eltérés megfelel-e annak, hogy csak a pontok közötti statisztikus ingadozás miatt különbözik.

$\chi^2$ -próba arra, hogy valami az elméleti gyakoriságnak megfelelően történt-e:

$$\mathbf{H}_0: p_1^{\text{elméleti}} = p_1^{\text{kísérleti}}, p_2^{\text{elméleti}} = p_2^{\text{kísérleti}}, \dots, p_N^{\text{elméleti}} = p_N^{\text{kísérleti}}$$

$\mathbf{H}_a$ : legalább egy egyenlőtlenség  $H_0$ -ban

Elvi megoldás:

$$\xi = \sum_{i=1}^N \frac{(y_i - e_i)^2}{e_i}, \text{ ahol } y_i = \text{az } i\text{-dik fajta eredmény megvalósulásainak száma, } e_i = kp_i, k = \sum_{i=1}^N y_i$$

ha  $\xi \leq \chi^2(\alpha, N-1)$ ,  $H_0$ -t elfogadjuk

ha  $\chi^2(\alpha, N-1) < \xi$ ,  $H_0$ -t elutasítjuk ( $H_a$ -t fogadjuk el)

A fenti séma bármi diszkrét függvényre tehető, arra is, ha két függvényt akarunk

összehasonlítani:  $g(x) \rightarrow g(x_i)$  és  $f(x) \rightarrow f(x_i)$

Megoldás EXCEL-lel: (új függvény neve: KHINÉGYZET.PRÓBA)

ha  $\alpha \leq \text{KHI.PRÓBA}(\text{adattömb}_{\text{tényleges}}; \text{adattömb}_{\text{várható}})$ ,  $H_0$ -t elfogadjuk

ha  $\text{KHI.PRÓBA}(\text{adattömb}_{\text{tényleges}}; \text{adattömb}_{\text{várható}}) < \alpha$ ,  $H_0$ -t elutasítjuk ( $H_a$ -t fogadjuk el)

### Feladatok

1) Állítson fel a várható értékekre és a szórásokra hipotéziseket és vizsgálja meg azokat statisztikai próbákkal a következő adatsorokra! Végezzen egymintás t-próbát is  $E_0 = 1,6$  és  $E_0 = 1,8$  értékekkel!

2,1	1,6
2,2	2,2
2,3	1,4
1,9	2,2
2,2	1,8
3,2	1,3

2) Az alábbi értékeket mérték ajkai iskolákban a beépített építőanyagok sugárzására (a dózisok dimenziómentesen vannak megadva). Modellezhető-e a mérés Poisson eloszlással?

dózis	gyakoriság
0	4
1	9
2	13
3	12
4	2
5	0

3)  $\text{AgNO}_3$  oldat vezetőképességére három hallgató az alábbiakat mérte ( $T=298 \text{ K}$ ,  $c=0,05 \text{ mol/dm}^3$ ). Elemezze statisztikai alapon a méréseket (várható értékek, szórások, konfidencia intervallumok, t- és F-próbák, egymintás t-próba, ha az elméleti érték  $E_0 = 115,2 \text{ cm}^2 \Omega^{-1} \text{ dm}^{-3}$ )! A konfidencia intervallumhoz az inverz t-eloszlás értékét az  $\text{INVERZ.T}(\alpha, N-1)$  függvénnyel kapja meg.

vezetőképesség $\text{cm}^2\Omega^{-1}\text{dm}^{-3}$ egységben		
1.hallgató	2.hallgató	3.hallgató
115,28	114,42	115,36
115,52	115,28	116,16
114,76	115,58	115,68
115,80	114,90	115,50

## VARIANCIAANALÍZIS - EGY TÉNYEZŐ SZERINTI OSZTÁLYOZÁS

Cél: A mért adatok különböző részekre oszthatóak: pl. más laborban mérték azokat, egy részük férfiakra/nőkre vonatkozik... Vajon van-e szignifikáns-e az eltérés a csoportok között?

Háttér:

$$SS_T = \sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 = SS_{csopbelül} + SS_{csopközött} = \sum_j \sum_i (x_{ij} - \bar{x}_{.j})^2 + \sum_j n(\bar{x}_{.j} - \bar{x}_{..})^2,$$

ahol "." a megfelelő indexre való átlagolást jelenti. A teljes varianciát két részre osztjuk: egy a csoportokon belüli és egy a csoportok átlagai közöttire. A megfelelő varianciák, ahol  $n$  az egy csoportban levő adatok száma,  $q$  a csoportok száma:  $SS_T/(nq-1)$ ,  $SS_{csopbelül}/q(n-1)$ ,  $SS_{csopközött}/(q-1)$ .

Hipotézis a csoportok várható értékeire:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_q$$

$H_a$ : legalább egy egyenlőtlenség  $H_0$ -ban

Elfogadjuk, ha kisebb, mint a megfelelő kritikus érték:

$$F = \frac{SS_{csopközött} / (q-1)}{SS_{csopbelül} / q(n-1)} < F_{q-1, q(n-1)}^{\alpha=0,05}$$

Megoldás EXCEL-lel: Adatok/Adatelemzés/Egytényezős varianciaelemzés

*Feladatok*

*Tej aflatoxin tartalmának mérése több laborban (betű = laborok jele)*

a	b	c	d	e	f	g
1,6	4,6	1,2	1,5	6	6,2	3,3
2,9	2,8	1,9	2,7	3,9	3,8	3,8
3,5	3	2,9	3,4	4,3	5,5	5,5
1,8	4,5	1,1	2	5,8	4,2	4,9
2,2	3,1	2,9	3,4	4	5,3	4,5

### Műszerek statisztikai ellenőrzése

A hallgatói laboratórium 5 pH-mérőjét a félév kezdete előtt ellenőrizték. A standard oldatból készülékenként 7-7 aliquot mintával mértek. Végezze el az adatok statisztikai analízisét (átlag, variancia, szórás, ANOVA, t- és F-próbák...). Az eredmények ismeretében tegyen javaslatot, melyik készüléke(ke)t kell újra beállítani (eltolódás a skálán) és melyik készüléke(ke)t kell javításra elküldeni (nagy véletlen hibával mér).

Mért adatok:

A	B	C	D	E
7,137	7,120	7,087	7,317	7,091
7,113	7,151	7,110	7,286	7,028
7,103	7,104	7,098	7,315	7,080
7,127	7,147	7,075	7,283	7,117
7,100	7,144	7,092	7,298	7,091
7,095	7,108	7,117	7,285	7,162
7,104	7,133	7,103	7,272	7,114

### MÁTRIXMŰVELETEK

Mátrix: érték és hely is számít:

összeadás  $C=A+B$ ,  $\leftrightarrow c_{ik}=a_{ik}+b_{ik}$

szorzás konstanssal:  $\text{const } A \leftrightarrow \text{const } a_{ik}$

mátrixok szorzása (sor-oszlop szorzás):  $C_{ln}=A_{lm}B_{mn} \leftrightarrow c_{ij} = \sum_{k=1}^m a_{ik} b_{kj}$

$$\text{egység mátrix: } I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

inverz mátrix:  $AA^{-1}=A^{-1}A=I$

$n \times n$ -es mátrix determinánsa:

$\det A = \sum_p (-1)^I a_{1k_1} a_{2k_2} \dots a_{nk_n}$ , az összes lehetséges olyan szorzat összege, ahol minden

sorból veszünk egy elemet, de az elemek más-más oszlopban vannak, ezeket összeszorozzuk és megszorozzuk +1-gyel van -1-gyel, attól függően, hogy páros, vagy páratlan oszlopserével hozható ez létre.

$\det A = 0$ , ha A egyik sorának az összes eleme = 0,

ha A egyik sora egy másik sor konstans szorosa

ha A egyik sora más sorok lineáris kombinációjával előállítható

ugyanaz az oszlopokra is vonatkozik

mátrixműveletek az EXCEL-ben:

MDETERM, MSZORZAT, INVERZ.MÁTRIX

(tömbfüggvény bevitele ctrl/shift/enter, lásd még a GYAKORISÁG függvényénél).

## LINEÁRIS REGRESSZIÓ

### Egyenes illesztése legkisebb négyzetek módszerével

$n$  darab  $x_i, y_i$  számpár esetén  $y=f(x)=ax+b$  egyenes illesztésének egyenletei:

minimalizálandó célfüggvény (szélsőérték):  $\sum (f(x_i) - y_i)^2$ ,  $a$  és  $b$  mint változók függvényében

$$a = \frac{\sum (x_i y_i - \bar{x} \bar{y})}{\sum (x_i x_i - \bar{x} \bar{x})} \text{ és } b = \bar{y} - a \bar{x}, \text{ ahol a felülvonás az } x_i \text{ és } y_i \text{ értékek átlagát jelenti}$$

reziduális (maradék):  $r_i = y_i - f(x_i)$ , ezeknek összege zérus

Origón átmenő egyenes esetén:  $y=f(x)=ax$  és  $a = \frac{\sum x_i y_i}{\sum x_i x_i}$

Legkisebb négyzetek módszerének grafikus szemléltetése

### Illesztés grafikonon

mérési pontok esetében, ha az  $y=f(x)$  függvény analitikus alakja vagy paraméterei nem ismertek, azok esetleges illesztése, valamint egy **adott  $x_0$  értékhez  $y_0$**  meghatározása grafikon segítségével:

grafikon készítése

az illesztendő görbe kijelölése

Trendvonal beszúrása

illesztendő görbe kiválasztása

illesztett görbe egyenletének kiírása, esetleges átmásolása, és a keresett  $x_0$  értéknél való kiszámítása

Feladatok:

### Molekulatömeg meghatározása a tökéletes gáztörvényből:

(Griffiths-Thomas: Fizikai kémiai számítások, 3.6. példa)

Minden gáz tökéletes gázként viselkedik végtelen kis nyomáson és kellően magas hőmérsékleten. Ha a nyomás függvényében a sűrűség/nyomás értékét ábrázoljuk, és 0 nyomásra extrapolálunk, a kapott tengelymetszet kapcsolatba hozható a gáz molekulatömegével.

Elmélet:

$$pV=nRT$$

R moláris gázállandó

$$n=m/M$$

$m$  a bemért tömeg, M a móltömeg

$$M=(mRT)/(pV)$$

$$m=\rho V \quad \rho \text{ a sűrűség}$$

$$M=(\rho/p)RT$$

Ismeretlen  $AH_3$  összetételű gázzal az alábbi értékeket mértük. Melyik az A elem?

Adatok:

$p$ (Pa)	$\rho/p$ (g/m <sup>3</sup> Pa)
101300	0,015111
75975	0,015076
50650	0,015042
25325	0,015010

$$M_H=1,008 \quad T=273K$$

$$M_A=(\text{tengelymetszet} \cdot 8,314 \cdot 273) - 3 \cdot 1,008$$

*Feladat: a fenti feladat megoldása máshogy is (TREND, METSZ, MEREDEKSÉG, LIN.ILL függvényekkel)*

Hűlési sebesség számítása I.: 10 másodpercenként lett leolvasva a rendszer hőmérséklete.

$t/s$	0	10	20	30	40	50	60	70	80
$T/K$	366	364	357	351	348	341	336	333	327

Mi lehet a hűlési sebesség dimenziója? Mi a függő és mi a független változó? Az illesztett egyenes alapján mennyi volt a rendszer hőmérséklete 33 s-kor? 120 másodperckor? A mérés megkezdése előtt 1 perccel!?

Hűlési sebesség számítása II., 10 fokként lett rögzítve az eltelt idő.

$t/s$	0	8	18	30	42	53	58	72	87
$T/K$	500	490	480	470	460	450	440	430	420

Mi a függő és mi a független változó? Hogyan illeszthető rá egyenes? Átmenjen-e az origón az illesztett egyenes? Műszeres mérési analógiák, kalibráció. Lehetőségek: a)  $x=t, y=T, v=a$  b)  $x=T, y=t, v=1/a$  ... Számítsuk ki az összes (jó és rossz) meredekséget! Mi az eredmény, ha egyre több középső pontot kihagyunk az illesztésből? Illeszthetjük-e az egyenest a  $t_{i+1}-t_i$  változásokra  $x_i=10$  K értékekkel?

### Inhomogén lineáris egyenletrendszer egyértelmű megoldása

Matematikai ismétlés:

Lineáris egyenletrendszer általános alakja:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \dots & \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n \end{aligned} \quad \leftrightarrow \underline{A}\underline{x}=\underline{b}$$

ahol  $A \in \mathbb{R}^{n \times n}$ ,  $\underline{x} \in \mathbb{R}^n$ ,  $\underline{b} \in \mathbb{R}^n$ , azaz

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}, \underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} \text{ és } \underline{b} = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix}.$$

Inhomogén a lineáris egyenletrendszer, ha legalább egy  $b_i \neq 0$ . Ha az összes  $b_i = 0$ , akkor homogénnek nevezzük, ezzel mi most nem foglalkozunk. Az inhomogén lineáris egyenletrendszer akkor oldható meg egyértelműen, ha  $\det A \neq 0$ .

Ha  $\det A = 0$ , akkor szingulárisnak nevezik a mátrixot. (hasonló fogalmak ugyanerre: rang, vektorok függetlensége)

*Feladat: piaci vásárlás példája (3 fajta gyümölcs-3 vásárló; 3 fajta gyümölcs+ zacskó-4 vásárló)*

EXCEL-lel: LIN.ILL függvény

### Túlhatározott lineáris egyenletrendszer megoldása

Matematikai ismételés: Több egymástól független sor ( $n$  darab), mint ahány ismeretlen ( $m$  darab). Az előzőhöz képest szerepcseré:  $a_{ij}$  igazából a  $j$ -dik független változó  $i$ -dik mérésben való értéke, amit korábban  $x_j$ -vel jelöltük, az most a meredekség, illetve az 1 együttható értékekhez tartozó érték a konstans tag. Több dimenziós egyenes illesztése: cél a meredekségek és a konstans tag meghatározása. Cél, hogy a számolt és mért eredményvektor négyzetösszege minimális legyen. Vagyis

$$\sum_{i=1}^n (b_i^{\text{mért}} - b_i^{\text{számolt}})^2 \text{ minimumát keressük.}$$

Levezethető megoldás:  $\underline{x} = (A^T A)^{-1} A^T \underline{b}$

Megoldás EXCEL-lel: LIN.ILL függvényvel, Eszközök/Adatelemzés/Regresszió

Eredmények értelmezése! illeszkedés jó, ha  $R^2$  érték közel van 1-hez

*Feladatok:*

*Piaci vásárlás példája (3 fajta gyümölcs, esetleg zacskó-5 vásárló)*

### Koncentráció meghatározása spektroszkópiai adatokból

Egy oldat különböző szerves anyagokat tartalmaz. A  $\lg\left(\frac{I_f}{I_0}\right) = -\sum_i \epsilon_i c_i l$  összefüggés alapján az A,

B, C, D és E anyagok koncentrációi öt különböző hullámhossznál történt mérés alapján meghatározhatóak. Az oldószer az adott hullámhosszoknál nem abszorbeál.

Az ismert moláris abszorpciós együtthatók:

	$\epsilon_A / (dm^3 \cdot mol^{-1} \cdot cm^{-1})$	$\epsilon_B / (dm^3 \cdot mol^{-1} \cdot cm^{-1})$	$\epsilon_C / (dm^3 \cdot mol^{-1} \cdot cm^{-1})$	$\epsilon_D / (dm^3 \cdot mol^{-1} \cdot cm^{-1})$	$\epsilon_E / (dm^3 \cdot mol^{-1} \cdot cm^{-1})$
$\lambda=300$ nm	114,3	10,1	2,0	26,7	56,3
$\lambda=400$ nm	3,0	89,1	4,2	22,1	19,8
$\lambda=500$ nm	10,0	9,7	160,1	30,1	2,0
$\lambda=600$ nm	64,5	5,6	20,1	230,4	11,4
$\lambda=700$ nm	19,4	4,5	8,7	10,8	132,3

A mért abszorbanciák:

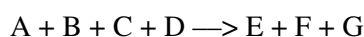
	$-\lg(I/I_0)$
$\lambda=300$ nm	0,269
$\lambda=400$ nm	0,197
$\lambda=500$ nm	0,331
$\lambda=600$ nm	0,297
$\lambda=700$ nm	0,231

A kűvetta vastagsága 1 cm.

(megoldás: 0,00138; 0,00163; 0,00175; 0,00065; 0,00132)

### Reakciósebességi állandó (k) meghatározása

Az alábbi bruttó egyenlettel leírható kémiai reakció sebességi állandóját keressük.



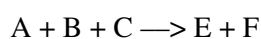
Mivel a reakció több lépésben megy végbe, ezért nem ismerjük a rendűségét sem. A mérést úgy végezzük, hogy bizonyos időközönként mintát veszünk az oldatból, és meghatározzuk az egyes komponensek koncentrációját. Ebből kiszámítjuk a koncentráció változásának a sebességét. A számítási eredményeket az alábbi táblázat tartalmazza:

Reakciósebesség	$c_A, (mol/dm^3)$	$c_B, (mol/dm^3)$	$c_C, (mol/dm^3)$	$c_D, (mol/dm^3)$
7,74E-04	2,3	0,9	1,9	2,2
5,90E-04	1,7	0,8	1,7	2,1
3,87E-04	1,4	0,7	1,4	1,8
2,60E-04	0,7	0,6	1,3	1,3
1,51E-04	0,3	0,5	1,0	1,2

Írja fel a lineáris egyenletrendszeret a  $v=k[A]^a[B]^b[C]^c[D]^d$  egyenlet logaritmizálásával és oldja meg! Mekkora  $k$  értéke? (megoldás: 2,84e-4)

### Reakciósebességi állandó (k) meghatározása többdimenziós egyenes illesztésével

Az alábbi bruttó egyenlettel leírható kémiai reakció sebességi állandóját keressük.



Mivel a reakció több lépésben megy végbe, ezért nem ismerjük a rendűségét sem. A mérést úgy végezzük, hogy bizonyos időközönként mintát veszünk az oldatból, és meghatározzuk az egyes komponensek koncentrációját. Ebből kiszámítjuk a koncentráció változásának a sebességét. A számítási eredményeket az alábbi táblázat tartalmazza:

Reakciósebesség	$c_A$ , (mol/dm <sup>3</sup> )	$c_B$ , (mol/dm <sup>3</sup> )	$c_C$ , (mol/dm <sup>3</sup> )
7,74E-04	0,27	0,15	0,36
5,90E-04	0,24	0,13	0,32
3,87E-04	0,20	0,10	0,28
2,82E-04	0,17	0,07	0,24
2,40E-04	0,16	0,07	0,23
2,20E-04	0,15	0,06	0,22
1,82E-04	0,14	0,06	0,20
1,51E-04	0,12	0,05	0,19

Írja fel a túlhatározott lineáris egyenletrendszer a  $v=k[A]^a[B]^b[C]^c$  egyenlet logaritmizálásával és oldja meg! Mekkora  $k$  értéke? (megoldás:0,0107)

#### Lineáris regresszió, paraméterek és megbízhatósági intervallumai

Az  $y$  mennyiség lineárisan függ az A, B, C anyagok koncentrációjától. Határozza meg lineáris regresszióval a három anyagra vonatkozó állandót ( $m_j$ -t, meredekséget) az alábbi adatsor alapján. Az illesztett egyenesnél a  $b$  konstans tag értéke eltérhet 0-tól. Adja meg a paraméterek megbízhatósági intervallumát is 95 %-os kétoldali konfidencia intervallummal az alábbi képlet alapján:  $m_j \pm t_{(n-p,\alpha)} * s_j$ ; A paraméterek szórásai ( $s_j$ -k) a négyzetgyökei az  $S_r^2 (X^T X)^{-1}$  mátrix diagonális elemeinek, ahol

$$S_r^2 = \frac{\sum_{i=1}^n (y_i^{\text{mért}} - y_i^{\text{illesztett}})^2}{n - p}$$

a reziduális szórásnégyzet,  $n$  a mérések száma,  $p$  a paraméterek száma.  $X$  a

független változók mátrixa a konstans taghoz tartozó 1-eseket tartalmazó oszloppal együtt. A  $t$ -eloszlás értékét közelítheti a normáleloszlás 95%-os kétoldali megbízhatósági értékének 1,96-os szorzójával. Mért adatok:

$c_A$	$c_B$	$c_C$	$y$
0,050	0,030	0,020	0,174
0,020	0,014	0,023	0,052
0,012	0,014	0,013	0,050
0,012	0,015	0,012	0,072
0,010	0,034	0,034	0,097
0,005	0,005	0,005	0,042
0,067	0,002	0,001	0,150



## PARAMÉTERBECSLÉS

(Libre Office Calc: nem megy, nincs benne minimalizáló és tetszőleges véletlenszámgeneráló)

### Tetszőlegesen generált, majd hibával torzított függvény paramétereinek becslése

A-B oszlop: kitöltés 10-10 tetszőleges számmal (pl. egész számok)

C oszlop:  $C1=1,2*\cos(A1)-2/B1$

D oszlop: standard normális eloszlású véletlen számok generálása Adatelemzés/Vélsz.generálás

E oszlop:  $E1=C1+0.05*D1$  (hibát generáltunk az adatokhoz)

F1;F2 p1 és p2 paraméterek kezdőértékei

G oszlop: C oszlop képlete, de 1,2 helyett \$F\$1 és 2 helyett \$F\$2

$F3=\text{SZUMXBÖLY2}(C1:C10;G1:G10)$

F3 minimumának megkeresése p1 és p2 függvényében Solverrel

Ugyanez E és G oszlopra

### Konszekutív kémiai reakció sebességi állandóinak meghatározása

Az  $A \xrightarrow{k_1} B \xrightarrow{k_2} C$  konszekutív kémiai reakció differenciálegyenlete megoldható analitikus módon.  $A(t=0)=1 \text{ mól/m}^3$ ,  $B(t=0)=0 \text{ mól/m}^3$ , és  $C(t=0)=0 \text{ mól/m}^3$  feltételek esetén:

$$C(t) = 1 - \frac{k_2}{k_2 - k_1} e^{-k_1 t} + \frac{k_1}{k_2 - k_1} e^{-k_2 t}$$

Paraméterbecsléssel határozza meg a következő szimulált adatokra a  $k_1$  és  $k_2$  sebességi állandókat. Az illesztést végezze el mind a pontos, mind a hibával terhelt adatokra. Kezdőértékként  $k_1=4$ -et és  $k_2=5$ -öt használjon.

$t(\text{s})$	$C(t) (\text{mól/m}^3)$ (pontos)	$C(t) (\text{mól/m}^3)$ (hibával terhelt)
0,5	0,2018	0,2042
1	0,4731	0,4025
2	0,7982	0,8534
3	0,9254	0,9095
4	0,9725	0,9733
5	0,9898	0,9072
6	0,9962	0,9740
7	0,9986	0,9877
8	0,9995	1,0233
9	0,9999	1,0370

### Ammónia van der Waals állandóinak becslése

*Elméleti háttér:*

Reális gázok leírását szolgálja a van der Waals egyenlet, melynek alakja:

$$\left( p + \frac{an^2}{V^2} \right) (V - nb) = nRT \quad (1).$$

Ebből a nyomást kifejezve:

$$p = \frac{nRT}{V - nb} - \frac{an^2}{V^2} \quad (2).$$

A van der Waals egyenlet alapján kapcsolat található az adott gáz kritikus hőmérséklete és nyomása, valamint az  $a$  és  $b$  állandók között. A számítások részletezése nélkül:

$$p_{kr} = \frac{a}{27b^2} \quad (3) \quad \text{és} \quad T_{kr} = \frac{8a}{27bR} \quad (4)$$

*Feladatok:*

1) A kritikus értékek és a 3-4. egyenletek alapján határozza meg az ammónia  $a$  és  $b$  van der Waals állandóját. A számolások megoldhatóak számológéppel vagy EXCEL-lel is. Az ammónia gáz kísérleti kritikus adatai:  $T_{kr} = 405 \text{ K}$ ,  $p_{kr} = 11,298 \cdot 10^6 \text{ Pa}$ .

2) A mérési adatok alapján nemlineáris paraméterillesztéssel is határozza meg a van der Waals egyenlet  $a$  és  $b$  paramétereit ammóniára (2. egyenlet felhasználásával; a kritikus értékből korábban számoltak alapján becsülje a paraméterek kezdőértékét). Adja meg az illesztett egyenlet alapján számolt nyomásokat is.

Ammónia gáz kísérleti móltérfogatai különböző nyomásokon 323,15 K hőmérsékleten:

$V_m \text{ (m}^3\text{/mol)}$	$p \text{ (Pa)}$
22,13E-3	1,202E5
5,110E-3	5,046E5
2,550E-3	9,697E5
1,456E-3	1,594E6

3) Számolja ki, hogy az első módon meghatározott állandókkal mekkora nyomásértékeket kap. Grafikonon és táblázatban ábrázolja a kétféle paraméterkészlettel kapott eredményeket, valamint a kísérleti eredményeket! Értékelje a van der Waals egyenlet megbízhatóságát!