Příloha

1. Statistické zpracování experimentálních výsledků

Skutečnou hodnotu měřené veličiny bychom získali jen naprosto dokonalým měřením. To však není možné, protože každé měření je zatížené chybami. Proto se zavádí tzv. konvenční hodnoty, což může být třeba hodnota zveřejněná referenční autoritou (např. Avogadrovo číslo), nebo teoreticky vypočítané číslo (např. obsah dusíku ve sloučenině C_5H_5N). Za výslednou hodnotu experimentálního měření pak považujeme tzv **„nejlepší odhad“** zvolené veličiny získaný měřením. Správnost měření je těsnost shody výsledku měření se skutečnou (resp. konvenční skutečnou) hodnotou. Číselným vyjádřením správnosti individuálního měření je jeho **chyba,** kterou mohou tvořit v podstatě dvě složky:

- **náhodná chyba** - výsledek náhodných jevů působících v okamžiku měření, který nelze ani předpovědět, ani zopakovat, ani eliminovat. Extrémním případem je tzv. hrubá chyba, způsobená lidským nebo přístrojovým selháním; hodnotu zatíženou hrubou chybou je třeba buď přímo nebo po testu na odlehlé výsledky ze souboru měření vyloučit.
- **systematická chyba** - konstantní nebo spojitě se měnící příspěvek k hodnotě měřené veličiny, který lze zcela nebo částečně korigovat např. novou kalibrací, použitím testovacích látek apod.

Přesnost měření je těsnost shody mezi nezávislými výsledky opakovaných měření, tzn. čím menší jsou náhodné chyby jednotlivých měření, tím větší je přesnost měření. Charakterizuje se pomocí **průměru a nejistoty**.

Při opakovaných měřeních veličiny *A* naměříme postupně hodnoty A_1 až A_n**. Za nejlepší odhadem veličiny A je obvykle považován aritmetický průměr** nalezených hodnot A_p

$$A_p = \frac{1}{n} \cdot \sum_{i=1}^{n} A_i \qquad (1.)$$

kde *n* je počet měření. Čím je *n* větší, tím se hodnota A_p více blíží tzv. střední hodnotě *A*, která se v nepřítomnosti systematické chyby shoduje se skutečnou hodnotou; to znamená, že průměr je nejlepším odhadem střední hodnoty. Jestliže odchylku každého měření od tohoto průměru označíme ε_i

$$\varepsilon_i = A_i - A_p \qquad (2.)$$

pak **standardní nejistota** je definována jako **odhad směrodatné odchylky** základního souboru vztahem

$$s = \sqrt{\frac{\sum_{i=1}^{n} \left(\varepsilon_i\right)^2}{n-1}} \qquad (3.)$$

kde výraz ve jmenovateli $n - 1 = \nu$ je tzv počet stupňů volnosti.

Směrodatná odchylka aritmetického průměru, jako míra nejistoty průměru způsobená náhodnými jevy se vypočítá takto:

$$s_{xp} = \frac{s}{\sqrt{n}} \qquad (4.)$$

Zatímco **aritmetický průměr** je tzv. bodový odhad střední hodnoty, pomocí směrodatné odchylky můžeme definovat její intervalový odhad; můžeme určit interval hodnot kolem průměru, ve kterém bude s určitou pravděpodobností ležet střední hodnota základního souboru. K tomu se využije **rozšířená nejistota** U , která se vypočítá ze standardní nejistoty, t.j. směrodatné odchylky, násobením koeficientem rozšíření k :

$$U = k \cdot s \tag{5.}$$

Pro velký počet měření (alespoň $n > 20$) veličiny, pro kterou platí normální Gausovo rozdělení nahodilých chyb, se nejčastěji používá hodnoty $k = 2$, což odpovídá *95%*-ní spolehlivosti měření, je možné použít i $k = 3$ k dosažení *99.5%*-ní spolehlivosti. Použití $k = 1$ odpovídá pouze *68,22%*-ní spolehlivosti.

Je-li počet provedených měření malý, je nutné vycházet ze Studentova rozdělení, konstanta k se označuje obvykle jako t a její hodnoty uvádí **TABULKA IX**.

Výsledek vždy uvádíme včetně koeficientu k nebo s uvedením spolehlivosti v měření v procentech. Například pokud používáme **rozšířenou nejistotu**:

- obsah celkového dusíku je (3.52 ± 0.14) g, kde rozšířená nejistota je vypočtena za použití koeficientu rozšíření k=2, což odpovídá hladině spolehlivosti *95%*.

Nebo pokud uvádíme přesnost výsledku za použití **standardní nejistoty**:

- obsah celkového dusíku je 3.52 g, se standardní nejistotou 0.07g a odpovídá 1 směrodatné odchylce[1]

DŮLEŽITÉ: Přestože při výpočtu využíváme obvykle nepřiměřeně vysoké přesnosti dané výpočetní technikou, **numerická hodnota presentovaného výsledku a jeho nejistota nemají být uváděny nadbytečným počtem číslic**. Hodnotu průměru uvádíme na stejný počet desetinných míst, jaký mají jednotlivé experimentální výsledky. Počet desetinných míst u výsledné nejistoty by měl být stejný (maximálně o jedno větší) jako u průměru.

Lineární a jednoduchá nelineární regrese

V praxi se setkáváme s případy, kdy měřená veličina y je funkcí jednoho nebo více parametrů. Častým případem je, opakované naměření, kdy získáme dvojice experimentálních hodnot y_i, x_i, kde $i = 1$, až n. Hodnoty mohou být svázány lineární závislostí, pak platí že:

$$y_i = a + b \cdot x_i + \varepsilon_i \tag{6.}$$

nebo nelineární funkcí f:

$$y_i = f(x_i, a_1, ..., a_m) + \varepsilon_i \tag{7.}$$

Kde a a b (respektivě a $a_1, ..., a_m$) jsou parametry lineární (nelineární) závislosti a ε_i jsou tzv. reziduály. Hodnoty parametrů získáme postupem, kdy měníme parametry tak, aby součet druhých mocnin reziduálů ε_i byl minimální; je to tzv. metoda nejmenších čtverců. Vzhledem k tomu, že jednotlivá měření jsou zatížena chybou, je třeba vyhodnotit větší počet dvojic měření a to minimálně tři na jeden optimalizovaný parametr.

[1] Při uvádění přesnosti se standardní nejistotou *k=1* se nemá používat způsobu psaní s „±“.

Příloha

Regresi závislosti je možné provést s výhodou za použití již vytvořeného software. Jednou z možností je použít tabulkový procesor MS EXCEL. Nejlépe je postupovat tak, že sestavíme tabulku se sloupci hodnot x_i a y_i . Do buněk dalšího sloupce zadáme instrukce k výpočtu hodnoty $a + b \cdot x_i$ respektive $f(x_i, a_1, ..., a_m)$ s použitím odhadů výchozích parametrů a dané hodnoty x_i. V dalších sloupcích počítáme odchylky ε_i mezi skutečnými hodnotami y_i a hodnotami y_i vypočtenými z odhadů parametrů. Do závěrečného sloupce doplníme hodnoty $(\varepsilon_i)^2$ a nakonec sestavíme účelovou funkci $\sum_i (\varepsilon_i)^2$. Tabulkový procesor EXCEL je vybaven nástrojem „řešitel“, který provede minimalizaci účelové funkce a nabídne optimalizované hodnoty parametrů.

POSTUP VÝPOČTU REGRESNÍCH CHARAKTERISTIK LINEÁRNÍ ZÁVISLOSTI:

1. Pro hodnoty x_i a y_i vypočítáme aritmetické průměry x_p a y_p (nezaokrouhlujeme!).

2. Hodnoty x_i a y_i centrujeme, tzn. od každé odečteme odpovídající průměr: $x_{ic} = x_i - x_p$, $y_{ic} = y_i - y_p$..

3. Vypočítáme hodnoty výrazů sestavených z centrovaných hodnot: $\sum (x_{ic})^2$, $\sum (y_{ic})^2$ a $\sum (y_{ic} \cdot x_{ic})$

4. Parametry *a* a *b* pak počítáme dle těchto vztahů:

$$b = \frac{\sum (y_{ic} \cdot x_{ic})}{\sum (x_{ic})^2} \qquad (8.)$$

$$a = y_p - b \cdot x_p \qquad (9.)$$

5. Z parametrů *a* a *b* a nezávisle proměnných hodnot x_i vypočítáme vyrovnané hodnoty Y_i a reziduály ε_i jako rozdíly *y*-hodnot naměřených a vypočtených

$$Y_i = a + b \cdot x_i \qquad (10.)$$

$$\varepsilon_\mathrm{i} = \mathrm{y_i} - \mathrm{Y_i} \qquad (11.)$$

6. Úspěšnost regresního modelu se testuje pomocí **směrodatné odchylky regrese** s_E a pomocí **korelačního koeficientu** *r*. Regrese je tím lepší, čím je r bližší číslu 1.

$$s_E = \sqrt{\frac{(y_i - Y_i)^2}{n-2}} \qquad (12.)$$

$$r = \frac{\sum (y_{ic} \cdot x_{ic})}{\sqrt{\sum x_{ic}^2 \cdot \sum y_{ic}^2}} \qquad (13.)$$

$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}} \qquad (14.)$$

7. Korelační koeficient *r* se musí testovat ve vztahu k počtu experimentálních dat. Obvykle očekáváme, že bude významný alespoň na *0.1%*-ní hladině spolehlivosti, t.j.

$\alpha = 0.001$. Je-li hodnota *t* vypočítaná z hodnoty *r* větší než kritická hodnota *t*-rozdělení pro odpovídající počet stupňů volnosti tj. $\nu = n - 2$, lze z *99.9%*-ní pravděpodobností usoudit, že odpovídající lineární závislost nevznikla náhodou.

8. Ze směrodatné odchylky regrese vypočítáme směrodatné odchylky parametrů *a* a *b*

$$s_a = s_E \cdot \sqrt{\frac{1}{n} + \frac{x_p^2}{\sum x_{ic}^2}} \qquad (15.)$$

$$s_b = s_E \cdot \sqrt{\frac{1}{\sum x_{ic}^2}} \qquad (16.)$$

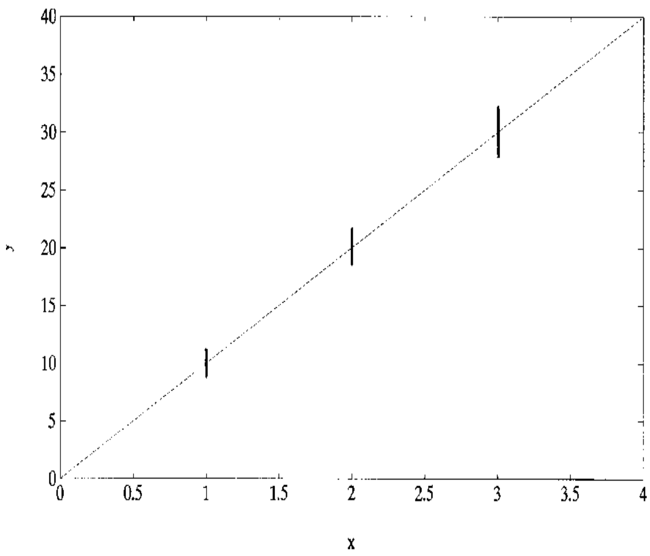
9. Otestujeme podíly a/s_a a b/s_b pomocí kritických hodnot t-rozdělení. Je-li vypočtená hodnota větší než tabelovaná pro $\alpha = 0.05$ a odpovídající počet stupňů volnosti, t.j. $\nu = n - 2$, pak se odpovídající parametr významně liší od nuly.

TABULKA IX: Kritické hodnoty t-rozdělení, kde ν je počet stupňů volnosti a α je hladina spolehlivosti (s údajem o spolehlivosti měření v procentech).

ν	$\alpha = 0.05$ (95%)	$\alpha = 0.001$ (99,9%)	ν	$\alpha = 0.05$ (95%)	$\alpha = 0.001$ (99,9%)
2	4.303	22.326	12	2.179	3.930
3	3.182	10.213	13	2.160	3.852
4	2.776	7.173	14	2.145	3.787
5	2.571	5.893	15	2.131	3.733
6	2.447	5.208	16	2.120	3.686
7	2.369	4.785	17	2.110	3.646
8	2.306	4.501	18	2.101	3.610
9	2.262	4.297	19	2.093	3.579
10	2.228	4.144	20	2.086	3.552
11	2.201	4.025	∞	1.960	3.090

Grafické znázornění regresní závisloti

Na osu úseček nanášíme nezávisle proměnnou x, jejíž hodnoty volíme, na osu pořadnic hodnotu její funkce y, kterou měříme. Nejistotu hodnot y_i vypočteme dle vztahů (1.) až (5.), kde A je nahrazeno y, a naznačíme úsečkou ve směru osy pořadnic kolem odpovídajícího bodu (viz **OBRÁZEK 1**).

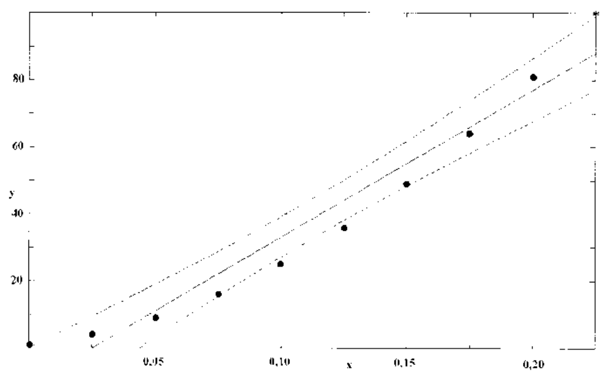


OBRÁZEK 1: Intervaly spolehlivosti hodnot *y* u regresní přímky pro *k=2*.

Pro jednotlivé vypočtené body na regresní přímce můžeme vyznačit pás spolehlivosti, t.j. pro každou hodnotu x_i vypočítáme s_{Yi} dle vztahu:

$$s_{Yi} = s_E \cdot \sqrt{\frac{1}{n} + \frac{\left(x_i - x_p\right)^2}{\sum \left(x_{ic}\right)^2}} \qquad (17.)$$

a po vynásobení odpovídající hodnotou *t*-rozdělení vyneseme získaný výsledek jako koncové body úsečky kolem bodu na regresní přímce. Při dostatečné hustotě bodů dostaneme kolem regresní přímky pás (viz **OBRÁZEK 2**), který se zužuje směrem ke středu grafu - nejužší je pro x_p. Se spolehlivostí, pro kterou bylo zvoleno t, každá další naměřená hodnota y padne do tohoto pásu.



OBRÁZEK 2: Regresní přímka s pásem spolehlivosti pro *k=2* (95%).

2. Přílohy

Osnova protokolu o vykonané laboratorní úloze

PROTOKOL O VYKONANÉ LABORATORNÍ ÚLOZE je základní dokument shrnující naměřená a vypočtená data. Hodnocení jeho obsahové i formální úrovně je jednou ze součástí celkového hodnocení posluchače. **Protokoly jsou zásadně vyžadovány vždy v následujícím laboratorním cvičení.** Pokud má protokol více stran nebo je provázen grafy či záznamem ze zapisovače, neoddělitelně vše spojíme v celek. Používáme papíry formátu A4.

HLAVIČKA A OBSAH PROTOKOLU MUSÍ OBSAHOVAT ÚDAJE V NÁSLEDUJÍCÍM VZORU[2]:

pořadové číslo[3]: 4/13	úloha: **5B[4]. ADSORPCE NA MEZIFÁZÍ KAPALINA-PLYN**	hodnocení:[5]
obor: 3.r. Ch-Ma[6]	posluchač(-ka): Markéta Neveselá	datum měření: 15. říjen 2003

PRINCIP: Podstata úlohy, základní teorie a vztahy, zejména ty, které je třeba k vyhodnocení naměřených hodnot.

POSTUP: Stručný popis konkrétního postupu a organizace práce při měření úlohy.

VÝSLEDKY: Tato nejrozsáhlejší část protokolu musí obsahovat vše, co je uvedeno v závěrečném odstavci úlohy s označením: **„ Protokol“**

Tabulky musí obsahovat v záhlaví sloupců či řádků symboly a jednotky prezentovaných dat. Používáme výhradně hlavní nebo vedlejší jednotky soustavy SI resp. jednotky od nich odvozené. Pokud tabulka obsahuje vypočtené hodnoty, pak pro vybraný řádek či sloupec tabulky (v tabulce vyznačíme) uvedeme pod tabulku vzorový výpočet (tj. obecný vztah, vzorové číselné dosazení a rozměrovou analýzu).

Grafy se zpracovávají s použitím grafických programů na PC. Každý graf musí mít název grafu a osy s popisem, tj. se symboly a jednotkami vynášených veličin. V grafech se vždy uvádí jak experimentální body (znázorněné zřetelnými grafickými symboly ve velikosti 3-5 *mm*) tak i křivky bez bodů odpovídající teoretickým závislostem. K bodům i křivkám musí být připojena v grafu legenda. Je vhodné připojit k experimentálním bodům úsečky vyjadřující interval spolehlivosti.

ZÁVĚR: Tato část protokolu obsahuje slovní odpověď na odstavec návodů **„ úkol“**, srovnání získaných veličin s jejich tabelovanými nebo očekávanými hodnotami, formulaci o přesnosti a správnosti naměřených veličin a vysvětlení případných příčin odchylek od očekávaných hodnot.

---

[2] Případné odchylky a rozsah práce stanoví vyučující.

[3] Posluchač uvádí pořadové číslo úlohy dle zveřejněného seznamu úloh lomené počtem úloh, které má absolvovat.

[4] Označení úlohy dle čísla kapitoly těchto skript.

[5] Místo pro záznamy učitele.

[6] Uvádí se ročník a zkratka oboru studia posluchače.