1969 — ACTA UNIVERSITATIS PALACKIANAE OLOMUCENSIS — TOM 30
FACULTAS RERUM NATURALIUM

Katedra experimentální fyziky a metodiky fyziky přírodovědecké fakulty
Vedoucí katedry: Prof. paed. dr. Josef Fuka

STATISTICKÉ ZPRACOVÁNÍ TESTŮ VĚDOMOSTÍ VE FYZICE

MILAN BEDNAŘÍK

(Předloženo dne 3. června 1968)

Úvod

V práci je uvedena jedna z metod matematické statistiky aplikované v didaktickém výzkumu, jehož cílem bylo objektivní ověření efektivnosti programovaného učení ve výuce fyziky na střední škole. Šlo o dílčí výzkum v oblasti vyučovacích metod, který se realizoval v rámci řešení dlouhodobého státního vědecko-výzkumného úkolu „Nové pojetí ve vyučování matematice a fyzice".

Výzkum byl organizován jako pedagogický experiment, který proběhl ve dvou etapách. První etapa v roce 1966 měla ráz orientačního pokusu ve čtyřech vyučovacích hodinách fyziky u žáků tří tříd (jedna třída tvořila experimentální soubor, dvě třídy soubor kontrolní) 1. ročníku SVVŠ v Olomouci. Zkušeností prvé etapy výzkumu, jejíž výsledky jsou zpracovány v pracích [1], [2], bylo použito ve druhé etapě v roce 1967, kdy byl pokus rozšířen na osm vyučovacích hodin a do čtrnácti tříd (experimentální soubor měl sedm tříd, kontrolní také sedm tříd) 1. ročníku SVVŠ v Olomouci, v Praze, v Opavě a v Moravské Třebové. Ve třídách experimentálního souboru se pracovalo metodami programovaného učení, ve třídách kontrolního souboru metodami tradiční výuky.

Žáci uvedených tříd představovali tzv. výběr, který reprezentoval sledovanou populaci, tj. základní soubor všech žáků daného věkového stupně studujících na daném typu školy. Reprezentativnost výběru [3] byla zajištěna jeho rozsahem (celkem 193 žáků experimentálního souboru a 217 žáků kontrolního souboru) a technikou výběru (použití náhodného výběru).

O základním souboru byla pak vyslovena tato pracovní hypotéza: Metody programovaného učení aplikované ve vyučování fyzice jsou vzhledem ke konečnému efektu učebního procesu přinejmenším tak účinné jako běžně používané metody klasické.

Základními prostředky výzkumu byly programovaný učební text a diagnostické testy. Programovaný učební text [4], vypracovaný na základě podrobné obsahové analýzy učiva zvoleného tématu a s ohledem na běžné principy programovaného učení [5], byl určen k samostatné práci studentů; zajišťoval souvislý výklad nového učiva a jeho prvotní upevňování prostřednictvím systému úkolů a kontrolních otázek. Opakování a prohlubování nových poznatků, jakož i celková organizace vyučovací hodiny, byla ponechána učiteli. Šlo tedy o určitou kombinaci forem samostatného studia s klasickými způsoby výuky.

Diagnostické testy neboli testy vědomostní, jejichž konstrukce se opírala jednak o předchozí rozbor fyzikálního učiva, jednak o poznatky psychologické [6] a statistické [7], měly poskytnout maximální objem informací o vyučovacích výsledcích ve sledovaných třídách: o množství a kvalitě vědomostí, o úrovni fyzikálního myšlení žáků, o jejich schopnosti aplikovat poznatky při řešení fyzikálních úloh apod.

Diagnostické testy jsou neocenitelným a dnešní pedagogickou vědou vysoce uznávaným nástrojem každého didaktického výzkumu. Mohou posloužit nejen k přesné indikaci skutečného stavu výuky ve třídě, ale také k posouzení účinnosti různých vyučovacích metod a ke zkoušení a hodnocení žákovských vědomostí. Proti jiným metodám mají řadu předností: jsou časově úsporné (v relativně krátkém čase umožňují prověřit vědomosti velkého počtu žáků), objektivní (kladou na všechny žáky stejné požadavky a používají stejných kriterií k měření jejich výkonů) a poskytují dostatek materiálu k dalšímu zpracování.

Ke statistickému zpracování bylo použito dvou druhů testů: vstupního a výstupního. Vstupní test (pretest) poskytl údaje o počátečním stavu žákovských vědomostí před použitím programu. Obsahoval 6 úkolů požadujících znalost fyzikálních veličin a jejich jednotek, neformální znalost některých fyzikálních vztahů, jednoduchou fyzikální úvahu a řešení příkladů. Výstupním testem (posttestem) byly získány údaje o konečném stavu vědomostí po absolvování tématu. Tvořilo ho 10 úkolů zaměřených na znalost, porozumění a použitelnost stěžejních poznatků celého tématu.

Pro vyhodnocení vstupního a výstupního testu byl zaveden jednotný bodovací systém s maximálním bodovým ziskem 50 bodů. Každý bod reprezentoval jednu správnou dílčí odpověď nebo jednu elementární myšlenkovou operaci. Testy, předtištěné pro každého žáka, byly zadávány učitelem za přítomnosti pracovníka výzkumu ve dvou variantách.

Statistická analýza testů

O přijetí či zamítnutí hypotézy o příslušné populaci rozhodovaly výsledky testů vědomostí získané od náhodně vybraného souboru žáků. Statistický rozbor těchto výsledků totiž umožňuje při optimální homogenizaci výběru predikovat vlastnosti celé populace.

Výsledek testu vědomostí jednoho náhodně vybraného žáka, kvantitativně vyjádřený bodovým ziskem, nazveme *náhodnou veličinou* X. Tuto veličinu, která je důsledkem působení řady náhodných faktorů, nelze u jednoho konkrétního žáka stanovit předem. Poněvadž však četnosti náhodné veličiny X vykazují často [8] *normální rozdělení*, dá se snadno určit pravděpodobnost jejího výskytu.

Hustota pravděpodobnosti normálního rozdělení pro $x(-\infty, \infty)$ je dána [9] *frekvenční funkcí*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \tag{1}$$

kde μ je *střední hodnota*, σ *směrodatná odchylka;* její druhou mocninu σ^2 nazýváme *rozptyl.* Parametr μ charakterizuje polohu normální distribuce (pro $x = \mu$ má funkce $f(x)$ maximum), parametr σ^2 variabilitu náhodné veličiny X (čím větší je σ^2, tím širší a plošší je graf funkce). Normální rozdělení s charakteristikami μ, σ^2 se značí $N(\mu, \sigma^2)$.

Pravděpodobnost, že náhodná veličina X leží v intervalu (x_1, x_2) je

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)\,\mathrm{d}x\,, \tag{2}$$

přičemž funkce

$$F(x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}x$$

je tzv. *distribuční funkce*. Pravděpodobnost vyjádřená rovnicí (2) odpovídá velikosti plochy omezené osou x, grafem frekvenční funkce (1) a přímkami $x = x_1$, $x = x_2$.

Frekvenční a distribuční funkce normálního rozdělení $N(\mu, \sigma^2)$ má v grafickém provedení různou polohu a tvar. Pro praktické potřeby se proto zavádí *standardizovaná náhodná veličina*

$$u = \frac{x - \mu}{\sigma}\,, \tag{4}$$

kterou se rozdělení $N(\mu, \sigma^2)$ transformuje na rozdělení $N(0,1)$, tj. na normální rozdělení se střední hodnotou $\mu = 0$ a rozptylem $\sigma^2 = 1$.

Příslušné frekvenční a distribuční funkce normálního rozdělení $N(0,1)$ jsou podrobně tabelovány [9].

Parametry normálního rozdělení $N(\mu, \sigma^2)$ charakterizují základní soubor, jsou tedy charakteristikami sledované populace. Poněvadž v praxi nelze pracovat s celým základním souborem, nahrazujeme populační charakteristiky jejich *odhady* získanými z výběrového souboru. Z teotie [8] vyplývá, že nejlepším odhadem populačního průměru μ je aritmetický průměr všech hodnot náhodné veličiny X neboli tzv. *výběrový průměr*

$$\bar{x} = \frac{1}{n} \sum_{1}^{n} n_i\, x_i\,, \tag{5}$$

kde n_i je absolutní četnost určité hodnoty x_i náhodné veličiny X, n je celkový počet členů výběru.

Nestranným odhadem populačního rozptylu σ^2 je veličina

$$s^2 = \frac{1}{n - 1} \sum_{1}^{n} n_i(x_i - x)^2, \tag{6}$$

která pro velká n je přibližně rovna *výběrovému rozptylu*

$$s = \frac{1}{n} \sum_{1}^{n} n_i\,(x_i - \bar{x})^2$$

Pro numerické výpočty upravujeme (6) na vztah

$$s^2 = \frac{1}{n-1}\left(\sum_{1}^{n} n_i x_i^2 - n\bar{x}^2\right). \tag{6'}$$

Odhadů populačních charakteristik používáme k *testování statistických hypotéz*, přičemž statistickou hypotézou rozumíme jakýkoliv výrok o populačních charakteristikách, tzn. také o frekvenční a distribuční funkci náhodné veličiny.

Při zpracování výsledků bylo použito: testu dobré shody, testu významnosti rozdílu mezi dvěma průměry a testu významnosti rozdílu mezi dvěma rozptyly.

Testem dobré shody byla testována hypotéza, že základní soubor, z něhož pochází náhodný výběr, má normální distribuci. Smysl testu, u něhož hodnoty x_i náhodné veličiny X sdružujeme vždy do několika tříd (skupin), spočívá v tom, že hodnotíme rozdíly mezi skutečnými experimentálně zjištěnými četnostmi n_i a hypotetickými četnostmi np_i teoretického rozložení. Veličina p_i je pravděpodobnost, že náhodná veličina nabude očekávané hodnoty. Testovacím kriteriem je podle [8] veličina

$$\chi^2 = \sum_{1}^{k} \frac{(n_i - np_i)^2}{np_i} = \sum_{1}^{k} \frac{n_i^2}{np_i} - n\,, \tag{7}$$

která má při dostatečně velkém n χ^2-rozdělení s $(k - 1)$ stupni volnosti; k je počet tříd. Hypotézu, že základní soubor má normální rozdělení na hladině významnosti α pak zamítáme, je-li

$$\chi^2 > \chi^2_\alpha\,,$$

kde χ^2_α je kritická hodnota χ^2-rozdělení s $(k - 1)$ stupni volnosti. Hodnoty χ^2_α určíme pro příslušnou hladinu významnosti z tabulek kritických hodnot χ^2-rozdělení [9]. Hladina významnosti α pak znamená jistý stupeň důvěry k našemu rozhodnutí o přijetí či zamítnutí dané hypotézy.

Test významnosti rozdílu mezi dvěma průměry byl proveden za účelem srovnání průměrných výsledků experimentální a kontrolní skupiny, a to jak při vstupní prověrce vědomostí (pretestu), tak i při výstupní prověrce (posttestu). Použito bylo t-testu pro dva nezávislé výběry [8].

Předpokládejme, že dva základní soubory mají normální rozdělení $N(\mu_1, \sigma_1^2)$ a $N(\mu_2, \sigma_2^2)$. Z každého odebereme po jednom výběru o rozsahu m a n s charakteristikami $\bar{x}$, s_x^2 a $\bar{y}$, s_y^2. Testujeme hypotézu, že průměry obou základních souborů jsou stejné, tj. $\mu_1 = \mu_2$, neboli $\mu_1 - \mu_2 = 0$ (takto formulovaný předpoklad nazýváme nulová hypotéza H_0), přičemž rozlišujeme dva případy:

$$\sigma_1^2 = \sigma_2^2,\ \sigma_1^2 \neq \sigma_2^2.$$

Je-li $\sigma_1^2 = \sigma_2^2$, je testovacím kriteriem veličina

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{s\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}}\,,$$

která má tzv. Studentovo t-rozdělení s $(m + n - 2)$ stupni volnosti. Totéž rozdělení má pak za předpokladu platnosti nulové hypotézy H_0 také testová charakteristika

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{m} + \frac{1}{n}}} \tag{8}$$

Veličina s je nestranný odhad populačního rozptylu $\sigma^2 = \sigma_1^2 = \sigma_2^2$, který je dán vztahem

$$s^2 = \frac{1}{m + n - 2}\left[\sum_1^m m_i x_i^2 - m\bar{x}^2 + \sum_1^n n_j y_j^2 - n\bar{y}^2\right]. \tag{9}$$

Vypočítanou hodnotu testovacího kriteria pak porovnáme s kritickou hodnotou $t_\alpha(f)$ Studentovy t-distribuce s $f = m + n - 2$ stupni volnosti, která je rovněž tabelována [9]. Je-li

$$|t| > t_\alpha(f),$$

pak hypotézu o rovnosti mezi dvěma průměry na zvolené hladině významnosti zamítáme a říkáme, že rozdíl mezi průměry je statisticky významný. Je-li

$$|t| \leqq t_\alpha(f),$$

nulovou hypotézu na dané hladině významnosti nezamítáme a tvrdíme, že sledovaný rozdíl není statisticky významný. Hladinou významnosti α rozumíme pravděpodobnost, že rozdíl průměrů $\bar{x} - \bar{y}$ překročí určitou hodnotu; pravděpodobnost, že tuto hodnotu nepřekročí, je pak $1 - \alpha$. V praxi volíme zpravidla $\alpha = 0{,}05$ nebo $\alpha = 0{,}01$.

Je-li $\sigma_1^2 \neq \sigma_2^2$, je postup poněkud složitější. Poněvadž tento případ při statistickém zpracování našich výsledků nenastal, neuvádíme ho.

K rozlišení dvou právě uvedených případů se používá *testu významnosti rozdílu mezi dvěma rozptyly*. Testujeme hypotézu, že $\sigma_1^2 = \sigma_2^2$. Testovacím kritériem je veličina

$$F = \frac{s_x^2}{s_y^2} \tag{10}$$

která má tzv. Fisher-Snedecorovo rozložení s $f_1 = m - 1$, $f_2 = n - 1$ stupni volnosti. Jestliže do čitatele zlomku (10) dosadíme větší odhad s^2 než do jmenovatele, pak hypotézu zamítáme na hladině významnosti α, je-li

$$F > F_{\frac{\alpha}{2}},$$

kde $F_{\frac{\alpha}{2}}$ je kritická hodnota F-rozložení, kterou najdeme v příslušných tabulkách. Hladina významnosti α vyjadřuje pravděpodobnost, že vypočítaná hodnota F padne mimo interval $(F_{\frac{\alpha}{2}}, F_{1-\frac{\alpha}{2}})$.

Ilustrujme uvedený postup na konkrétním případu diagnostického testu varianty A. Tabulka I ukazuje výsledky vstupního testu vědomostí varianty A, kterou absolvovalo $m = 100$ žáků experimentálního souboru a $n = 108$ žáků

Tabulka I. Vstupní test — varianta A

Experimentální soubor				Kontrolní soubor			
x_i	m_i	$m_i x_i$	$m_i x_i^2$	y_j	n_j	$n_j y_j$	$n_j y_j^2$
10	1	10	100	10	—	—	—
11	—	—	—	11	1	11	121
12	1	12	144	12	2	24	288
13	—	—	—	13	3	39	507
14	—	—	—	14	6	84	1 176
15	3	45	675	15	1	15	225
16	5	80	1 280	16	1	16	256
17	5	85	1 445	17	3	51	867
18	5	90	1 620	18	11	198	3 564
19	4	76	1 444	19	6	114	2 166
20	5	100	2 000	20	5	100	2 000
21	7	147	3 087	21	11	231	4 851
22	11	242	5 324	22	7	154	3 388
23	3	69	1 587	23	12	276	6 348
24	6	144	3 456	24	5	120	2 880
25	4	100	2 500	25	5	125	3 125
26	6	156	4 056	26	6	156	4 056
27	4	108	2 916	27	4	108	2 916
28	6	168	4 704	28	4	112	3 136
29	3	87	2 523	29	3	87	2 523
30	—	—	—	30	3	90	2 700
31	5	155	4 805	31	2	62	1 922
32	2	64	2 048	32	1	32	1 024
33	3	99	3 267	33	—	—	—
34	1	34	1 156	34	—	—	—
35	1	35	1 225	35	2	70	2 450
36	3	108	3 888	36	1	36	1 296
37	3	111	4 107	37	—	—	—
38	1	38	1 444	38	2	76	2 888
39	1	39	1 521	39	—	—	—
40	1	40	1 600	40	1	40	1 600
Σ	100	2 442	63 922	Σ	108	2 427	58 273

Experimentální soubor:

$$\bar{x} = \frac{2442}{100} = 24{,}42$$

$$\bar{x}^2 \doteq 596{,}34$$

$$m\bar{x}^2 \doteq 59\,634$$

$$s_x^2 = \frac{63\,922 - 59\,634}{99} \doteq 43{,}31$$

$$s_x \doteq 6{,}58$$

Kontrolní soubor:

$$\bar{y} = \frac{2427}{108} \doteq 22{,}47$$

$$\bar{y}^2 \doteq 505{,}00$$

$$n\bar{y}^2 \doteq 54\,540$$

$$s_y^2 = \frac{58\,273 - 54\,540}{107} \doteq 34{,}89$$

$$s_y \doteq 5{,}91$$

kontrolního souboru. V prvním sloupci tabulky jsou hodnoty x_i (resp. y_j) náhodné veličiny X (resp. Y), ve druhém sloupci jejich absolutní četnosti m_i (resp. n_j), třetí a čtvrtý sloupec obsahuje pomocné hodnoty použité k výpočtu odhadů populačních charakteristik $\bar{x}$ a s_x^2 (resp. $\bar{y}$ a s_y^2) podle vztahů (5) a (6').

Údaje tabulky II slouží k provedení testu dobré shody u experimentálního souboru. Hodnoty x_i náhodné veličiny X jsou zde sdruženy do $k = 10$ tříd, přičemž každá třída obsahuje interval tří hodnot; m_i jsou absolutní četnosti

Tabulka II

x_i	m_i	m_i^2	u_i	p_i	$m\, p_i$	$m_i^2/m\, p_i$
10–12	2	4	−1,89	0,029 38	2,938	1,361 4
13–15	3	9	−1,43	0,046 98	4,968	1,915 7
16–18	15	225	−0,97	0,089 66	8,966	25,094 8
19–21	16	256	−0,52	0,135 51	13,551	18,891 5
22–24	20	400	−0,064	0,172 95	17,295	23,128 0
25–27	14	196	0,39	0,177 25	17,725	11,057 8
28–30	9	81	0,85	0,150 61	15,061	5,378 1
31–33	10	100	1,304	0,101 54	10,154	9,848 3
34–36	5	25	1,76	0,056 92	5,692	4,392 1
37–40	6	36	2,37	0,030 31	3,031	11,877 2
						112,944 9

náhodné veličiny v příslušné třídě. Číslo u_i je hodnota standardizované náhodné veličiny dané vztahem (4), v němž jsou charakteristiky μ, σ^2 nahrazeny odhady $\bar{x}$, s_x^2. Pravděpodobnost p_i jednotlivých třídních intervalů určíme podle vztahu

$$p_i = \Phi(u_i) - \Phi(u_{i-1}),$$

kde $\Phi(u)$ je distribuční funkce normální distribuce $N(0,1)$; hodnoty této funkce hledáme opět v [9].

Podle (7) je

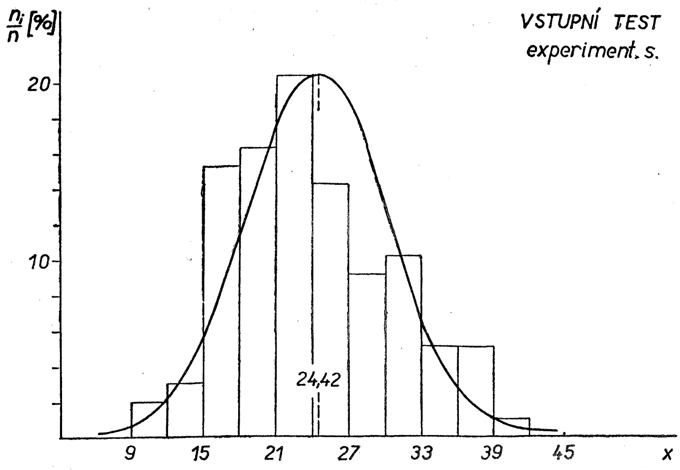
$$\chi^2 = 112{,}9 - 100 = 12{,}9.$$

Zvolíme hladinu významnosti $\alpha = 0{,}01$, určíme počet stupňů volnosti $f = (k - 1) - 2 = 7$ (za každý odhad populační charakteristiky snižujeme počet stupňů volnosti o 1; v našem případě o 2) a vyhledáme v tabulkách

$$\chi^2_{0,01}\,(7) \doteq 18{,}5\,.$$

Poněvadž je $\chi^2 < \chi^2(f)$, hypotézu o normálním rozdělení základního souboru na hladině významnosti $\alpha = 0{,}01$ nezamítáme. K analogickému výsledku se došlo testem dobré shody u kontrolního souboru. Hypotéza, že základní soubor, z něhož byla vybrána experimentální i kontrolní skupina, vykazuje normální rozložení, byla tedy přijata.

Grafické znázornění skupinového rozložení četností náhodné veličiny u experimentální skupiny je provedeno na obr. 1. Šířka každého sloupce histogramu představuje třídní interval, jeho výška relativní četnost náhodné veličiny v tomto



Obr. 1

intervalu. Histogramem je proložena teoretická křivka frekvenční funkce s maximem pro hodnotu, odpovídající odhadu populačního průměru.

K realizaci testu významnosti rozdílu mezi průměry a rozptyly obou souborů použijeme konečných výsledků tabulky I:

$$x = 24{,}42 \qquad y = 22{,}47$$
$$s_x^2 = 43{,}31 \qquad s_y^2 = 34{,}89$$
$$s_x = 6{,}58 \qquad s_y = 5{,}91$$

Nejdříve testujeme hypotézu o rovnosti dvou rozptylů. Hodnota testovacího kriteria podle (10) je

$$F = \frac{43{,}31}{34{,}89} \doteq 1{,}24$$

Pro hladinu významnosti $\alpha = 0{,}01$ a pro počet stupňů volnosti $f_1 = m - 1 = 99$, $f_2 = n - 1 = 107$ je podle tabulek kritická hodnota

$$F_{0{,}005}\,(99,\,107) \doteq 1{,}70$$

Poněvadž $F < F_{\frac{\alpha}{2}}\,(f_1, f_2)$, hypotézu nezamítáme a tvrdíme, že rozdíl mezi rozptyly není statisticky významný.

Nyní testujeme hypotézu o rovnosti dvou průměrů pro $\sigma_1^2 = \sigma_2^2$.

Podle (9) určíme nestranný odhad

$$s^2 = \frac{63922 - 59634 + 58273 - 54540}{100 + 108 - 2} \doteq 38{,}94$$

a odtud

$$s \doteq 6{,}24\,.$$

Podle (8) vypočítáme testovou charakteristiku

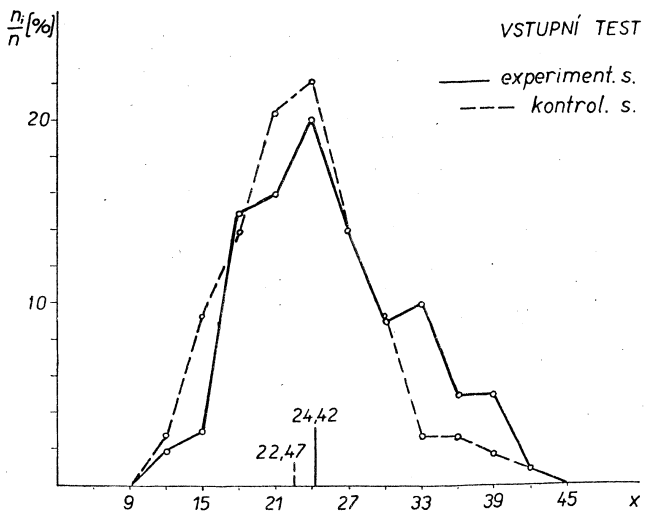
$$t = \frac{24{,}42 - 22{,}47}{6{,}24\sqrt{\frac{1}{100} + \frac{1}{108}}} \doteq 2{,}25 .$$

Z tabulek stanovíme kritickou hodnotu

$$t_{0,01}\,(100 + 108 - 2) \doteq 2{,}60$$

Poněvadž $|t| < t_\alpha(f)$, nulovou hypotézu nezamítáme a konstatujeme, že ani rozdíl mezi sledovanými průměry není na hladině významnosti $\alpha = 0{,}01$ signifikantní.

Z uvedených výsledků vyplývá, že náhodné výběry experimentální a kontrolní skupiny byly pořízeny z téhož základního souboru, který vykazuje normální rozdělení náhodné veličiny X, a že rozdíly mezi průměry obou výběrů nepovažujeme ze statistického hlediska za významné. Tento závěr potvrzuje také spojnicový diagram skupinového rozložení relativních četností vyznačený pro oba soubory na obr. 2. Poloha i tvar obou polygonů četností ukazují na celkovou shodu daných souborů.



Obr. 2

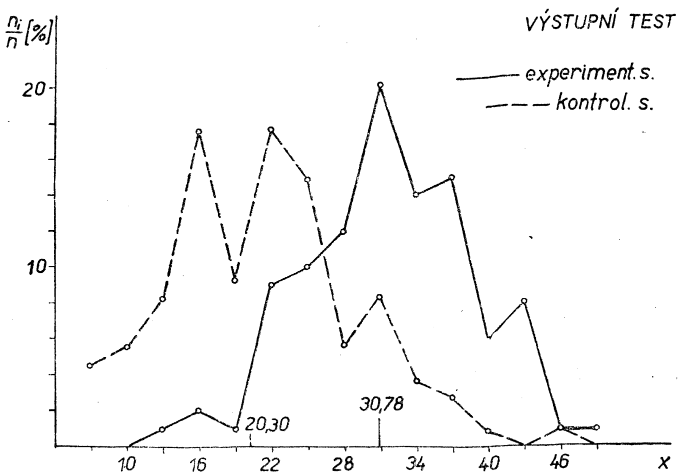
Výsledky výstupního testu vědomostí uvádí tabulka III. Lze z ní snadno vyčíst, že rozdělení náhodné veličiny je u obou sledovaných souborů nápadně odlišné. V experimentálním souboru jsou vyšší hodnoty četností zřetelně posunuty k vyšším hodnotám náhodné veličiny, což má za následek také vyšší číselnou hodnotu průměru $\bar{x}$. Ještě názorněji ilustruje tuto skutečnost umístění příslušných polygonů četností na obr. 3, z něhož je také zřejmý větší rozptyl a menší homogenita kontrolního souboru.

Tabulka III Výstupní test — varianta A

Experimentální soubor				Kontrolní soubor			
x_i	m_i	$m_i x_i$	$m_i x_i^2$	y_j	n_j	$n_j y_j$	$n_j y_j^2$
5	—	—	—	5	1	5	25
6	—	—	—	6	2	12	72
7	—	—	—	7	2	14	98
8	—	—	—	8	1	8	64
9	—	—	—	9	2	18	162
10	—	—	—	10	3	30	300
11	—	—	—	11	2	22	242
12	1	12	144	12	4	48	576
13	—	—	—	13	3	39	507
14	—	—	—	14	8	112	1 568
15	1	15	225	15	6	90	1 350
16	1	16	256	16	5	80	1 280
17	—	—	—	17	3	51	867
18	1	18	324	18	6	108	1 944
19	—	—	—	19	1	19	361
20	2	40	800	20	7	140	2 800
21	1	21	441	21	4	84	1 764
22	6	132	2 904	22	8	176	3 872
23	4	92	2 116	23	7	161	3 703
24	3	72	1 728	24	6	144	3 456
25	3	75	1 875	25	3	75	1 875
26	4	104	2 704	26	2	52	1 352
27	7	189	5 103	27	1	27	729
28	1	28	784	28	3	84	2 352
29	8	232	6 728	29	1	29	841
30	5	150	4 500	30	4	120	3 600
31	7	217	6 727	31	4	124	3 844
32	5	160	5 120	32	1	32	1 024
33	6	198	6 534	33	2	66	2 178
34	3	102	3 468	34	1	34	1 156
35	3	105	3 675	35	2	70	2 450
36	5	180	6 480	36	1	36	1 296
37	7	259	9 583	37	—	—	—
38	4	152	5 776	38	—	—	—
39	2	78	3 042	39	1	39	1 521
40	—	—	—	40	—	—	—
41	1	41	1 681	41	—	—	—
42	3	126	5 292	42	—	—	—
43	4	172	7 396	43	—	—	—
44	—	—	—	44	1	44	1 936
45	1	45	2 025	45	—	—	—
46	—	—	—	46	—	—	—
47	1	47	2 209	47	—	—	—
Σ	100	3 078	99 640	Σ	108	2 193	51 165
$\bar{x} = \frac{3078}{100} = 30{,}78$				$\bar{y} = \frac{2193}{108} \doteq 20{,}30$			

Pokračování tabulky III

$\bar{x}^2 \doteq 947{,}41$	$\bar{y}^2 \doteq 412{,}09$
$m\bar{x}^2 \doteq 94\,741$	$n\bar{y}^2 \doteq 44\,506$
$s_x^2 = \dfrac{99\,640 - 94\,741}{99} \doteq 49{,}48$	$s_y^2 = \dfrac{51\,165 - 44\,506}{107} \doteq 62{,}23$
$s_x \doteq 7{,}03$	$s_y \doteq 7{,}88$



Obr. 3

Provedeme statistický rozbor výsledků. Odhady populačních charakteristik jsou:

$$\bar{x} = 30{,}78 \qquad \bar{y} = 20{,}30$$
$$s_x^2 = 49{,}48 \qquad s_y^2 = 62{,}23$$
$$s_x = 7{,}03 \qquad s_y = 7{,}88$$

Testováním hypotézy o rovnosti dvou rozptylů zjistíme, že mezi rozptyly není významných rozdílů. Test hypotézy o rovnosti průměrů pro $\sigma_1^2 = \sigma_2^2$ při stejné hladině významnosti $\alpha = 0{,}01$ dává postupně tyto výsledky:

$$s^2 = 56{,}11, \quad s \doteq 7{,}49\,,$$
$$t = 10{,}04\,,$$
$$t_{0,01}\,(206) \doteq 2{,}60\,.$$

Z nerovnosti $|t| > t_\alpha(f)$ vyplývá, že nulovou hypotézu o rovnosti průměrů na dané hladině významnosti tentokráte zamítáme, což znamená, že rozdíl mezi sledovanými průměry je statisticky významný. Průměr $\bar{x} = 30{,}78$, kterým je kvantitativně vyjádřen výsledek učení experimentálního souboru, je významně vyšší než průměr $\bar{y} = 20{,}30$, vyjadřující výsledek učení souboru kontrolního.

Zajímavé jsou také vzájemné relace výběrových průměrů mezi vstupním a výstupním testem pro každý soubor. V experimentálním souboru jsou to hodnoty

$$\bar{x}_{\text{pretest}} = 24{,}42$$
$$\bar{x}_{\text{posttest}} = 30{,}78$$

a v kontrolním souboru

$$\bar{y}_{\text{pretest}} = 22{,}47$$
$$\bar{y}_{\text{posttest}} = 20{,}30$$

V prvém případě dochází ke zvýšení výkonu žáků (hodnota rozdílu $\bar{x}_{\text{posttest}} - \bar{x}_{\text{pretest}} = 6{,}36$), v druhém případě k jeho snížení ($\bar{y}_{\text{posttest}} - \bar{y}_{\text{pretest}} = -2{,}17$).

Analogické výsledky prokázala statistická analýza diagnostických testů varianty B.

Závěr

Pracovní hypotéza vyslovená při zahájení výzkumu se potvrdila. Komparační statistická měření výběrových průměrů, kterými byla kvantifikována soustava žákovských vědomostí na počátku a na konci experimentu, prokázala vyšší hladinu vyučovacích výsledků v experimentálním souboru. Vztáhneme-li tuto skutečnost na použité vyučovací metody, jsme oprávněni rezultovat, že metody programovaného učení byly v daném případě efektivnější nežli používané metody klasické.

LITERATURA

[1] *Bednařík, M.:* Programmed Instruction in Physics. – Programmed Learning, April 1967, s. 113–120.
[2] *Bednařík, M.:* Pokus o zařazení programovaného textu do hodin fyziky na SVVŠ. – Fyzika ve škole *6*, 1968, č. 7, s. 323–331.
[3] *Lindquist, E.:* Statistická analýza v pedagogickém výzkumu. – KPÚ, Praha 1965.
[4] *Bednařík, M.:* Gravitační pole. Programovaný učební text pro 1. ročník SVVŠ. Olomouc 1967.
[5] *Tollingerová, D.:* Teorie programovaného učení. In: Programované učení. – SPN, Praha 1966.
[6] *Meili, R.:* Lehrbuch der Psychologischen Diagnostik, Bern 1958.
[7] *Mittenecker, E.:* Planung und statistische Auswertung von Experimenten, Wien 1963.
[8] *Hampejsová, O., Břicháček, V.:* Statistické metody pro psychology, část I. – SPN, Praha 1964.
[9] *Janko, J.:* Statistické tabulky. – ČSAV, Praha 1958.

SUMMARY

STATISTICAL EVALUATION OF THE DIDACTIC TESTS IN PHYSICS

MILAN BEDNAŘÍK

The article deals with the statistical calculation of the results of the didactic research that was realized in the years 1966 and 1967 in 14 classes of the first form (aged 15—16) of the secondary schools. The aim of this research was to prove the effectiveness of the methods of programmed learning on the concrete physics subject-matter. By testing of knowledge the learning results of two accidentally selected groups of pupils were compared: the experimental group (7 classes with 193 pupils) was taught by means of a programmed textbook, the control group (7 classes with 217 pupils) was taught by teachers using the tradicional ways. The results of didactic tests are evaluated by means of the objective methods of mathemathical statistics. The testing of statistical hypotheses (the testing of the goodness of fit, the significance of the difference between two means, the significance of the difference between two variances) leads to the conclusion that the differences in knowledge of pupils after finishing the given subject-matter are significant in favour of the experimental group.

ZUSAMMENFASSUNG

STATISTISCHE AUSWERTUNG DER DIAGNOSTISCHEN TESTS IM PHYSIKUNTERRICHT

MILAN BEDNAŘÍK

Der Artikel befasst sich mit der statistischen Auswertung der Ergebnisse einer didaktischen Untersuchung, die in Jahren 1966 und 1967 in 14 Klassen des 1. Jahrganges der Oberschule verlief. Das Ziel dieser Untersuchung war die Wirksamkeit der Methoden des programmierten Unterrichts am konkreten Physiklehrstoff zu beweisen. Mit Hilfe von diagnostischen Testen wurden die Lehrergebnisse von zwei Zufallsstudentengruppen verglichen: einer experimentellen Gruppe (7 Klassen mit 193 Schülern), die mit dem Programmtext arbeitete, und einer Kontrollgruppe (7 Klassen mit 217 Schüllern), die mit üblichen traditionellen Methoden unterrichtet wurde. Die Testergebnisse wurden mit objektiven Methoden der statistischen Analyse bearbeitet. Das Testieren der statistischen Hypothesen stellt heraus, dass die Unterschiede in Schülerkenntnissen zwischen den beiden Gruppen nach Absolvieren des gegebenen Lehrstoffes signifikant zugunsten der experimentellen Gruppe ausfallen.

РЕЗЮМЕ

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ДИАГНОСТИЧЕСКИХ ТЕСТОВ В ФИЗИКЕ

МИЛАН БЕДНАРЖИК

В статье говорится о статистическом обсуждении результатов дидактического исследования, осуществившегося в 1966—67 гг. в 14 классах первого курса средней школы. Целью исследования было доказать эффективность методов программированного обучения на конкретном учебном материале по физике. При помощи диагностических тестов сравнивались учебные результаты двух случайно избранных групп учеников: экспериментальной группы (7 классов с 193 учениками), работающей с программированным учебным текстом, и контрольной группы (7 классов с 217 учениками), в которой преподавалось обыкновенным способом традиционного обучения. Результаты диагностических тестов обработаны объективными методами статистического анализа. Применением тестов статистических гипотез приходится к выводу, что различия в знаниях учеников после пройдения данного учебного материала между обеими группами с точки зрения статистики значительны в пользу экспериментальной группы.