Pokroky matematiky, fyziky a astronomie

Milan Bednařík; Miroslava Široká

Statická analýza výsledků písemné kontrolní práce z fyziky

Pokroky matematiky, fyziky a astronomie, Vol. 15 (1970), No. 3-4, 180--193

Persistent URL: http://dml.cz/dmlcz/139137

Literatura

[1] Fu: *Konference o vzdělání učitelů fyziky I. a II. cyklu.* FvŠ V. 1966/67. 233.
[2] Kašpar E. *Konference JČMF o vzdělávání učitelů fyziky.* FMFA. XII. 1967. 366.
[3] Kašpar E. *Základní problémy v přípravě učitelů fyziky.* FvŠ V. 1966/67. 363.
[4] Bělař A. *Otázky přípravy učitelů fyziky I. cyklu škol.* FvŠ VI. 1967/68. 20.
[5] Fuka J. *K přípravě učitelů fyziky na vysokých školách.* FvŠ.
[6] Kašpar E. *Problémy spojeného studia fyziky učitelů a neučitelů.* PMFA X. 1965. 82.

STATISTICKÁ ANALÝZA VÝSLEDKŮ PÍSEMNÉ KONTROLNÍ PRÁCE Z FYZIKY

Milan Bednařík, Miroslava Široká, Olomouc

Systematická kontrola vědomostí subjektu v procesu učení a adekvátní hodnocení výsledků této kontroly jsou nedílnou součástí výuky u většiny vyučovacích předmětů. Poskytují soubor nezbytných informací o reálném stavu výuky, zajišťují prostřednictvím činnosti zpětné vazby potřebnou kontinuitu procesu osvojování nových poznatků se soustavou poznatků dříve osvojených, vytvářejí základní podklady pro školní klasifikaci a v případě dostatečně velkého rozsahu i provedení některých přesnějších analýz pomocí metod matematické statistiky. Pro samotný subjekt učení jsou pak důležitým faktorem motivačním a prostředkem výchovným.

Kontrola a hodnocení úrovně vědomostí se provádí v pedagogické praxi různými způsoby. Jedním ze základních způsobů jsou písemné kontrolní zkoušky. Předmětem tohoto příspěvku je statistický rozbor některých výsledků získaných vyhodnocením písemné kontrolní práce z fyziky, která byla zadána skupině 127 posluchačů, přijatých ve studijním roce 1968/69 do prvního ročníku přírodovědecké fakulty UP v Olomouci. Skupinu tvořilo 58 posluchačů učitelského studia (28 matematika-fyzika, 14 fyzika-chemie, 16 matematika-deskriptivní geometrie) a 69 posluchačů studia odborného (12 jemná mechanika a optika, 16 analytická chemie, 41 numerická matematika). Chlapců bylo celkem 66, dívek 61. Všichni tito posluchači mají v prvním ročníku studia čtyřhodinovou nebo tříhodinovou přednášku z fyziky a dvouhodinové cvičení týdně; písemná práce jim byla zadána v první hodině cvičení z fyziky.

Písemná práce, jejíž obsah tvořilo základní učivo středoškolské mechaniky, se skládala ze dvou testů. Prvním testem byla prověřována především znalost jednotek mechanických veličin, druhým testem schopnost studentů aplikovat znalosti z mechaniky při řešení jednoduchých fyzikálních úloh a problémů. Z tohoto hlediska můžeme považovat výsledky prvního testu za míru formálních vědomostí posluchačů, výsledky druhého testu za míru znalostí neformálních. Úplné zadání obou testů a jejich diagnostický rozbor je uveden v práci [1].

Statistická analýza testů, která je zpracována v tomto článku, sledovala tyto hlavní úkoly:

a) vytvořit normu pro objektivní klasifikaci studentů ve cvičení z fyziky,

b) zjistit, zda existuje významný rozdíl ve znalostech chlapců a dívek,

c) určit stupeň korelace mezi výsledky prvého a druhého testu, tj. těsnost vztahu mezi formálními a neformálními znalostmi studentů.

POUŽITÉ PROSTŘEDKY STATISTICKÉ ANALÝZY

Statistické zpracování testů vědomostí předpokládá jednoznačné kvantitativní vyjádření jejich výsledků. V našem případě jsme zavedli zcela jednoduchý bodovací systém, podle něhož byl pro každého posluchače za každou správnou odpověď přiřazen jeden bod, a to bez ohledu na významnost nebo obtížnost otázky.

Nechť bodový zisk jednoho studenta, náhodně vybraného z daného souboru, představuje tzv. *kvantitativní náhodnou veličinu* X. Vzhledem k zavedenému bodovacímu systému nabývá tato náhodná veličina pouze diskrétních hodnot $x = 0, 1, 2, \ldots, k$, kde k je počet otázek testu. Počet studentů, jejichž bodový zisk je roven určité hodnotě x, označíme n_x a nazveme jej *absolutní četností* náhodné veličiny X. Počet všech studentů, kteří se podrobili zkoušce, označíme n. Je známo, že rozdělení četností náhodné veličiny lze v praxi často modelovat tzv. *normálním rozdělením* neboli *normální distribucí* [2]. Hustota pravděpodobnosti této distribuce je dána *frekvenční funkcí*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$

kde μ je *střední hodnota* měřené veličiny, σ *směrodatná odchylka* a σ^2 *rozptyl* neboli *variance*. Normální rozdělení s těmito parametry, které charakterizují polohu a tvar grafického vyjádření frekvenční funkce (Gaussovy křivky), označujeme $N(\mu, \sigma^2)$.

Pravděpodobnost, že hodnota x náhodné veličiny leží v intervalu (x_1, x_2), je

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

přičemž funkci

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

nazýváme *distribuční funkcí.* Poněvadž frekvenční a distribuční funkce normálního rozdělení $N(\mu, \sigma^2)$ jsou pro praktické použití nepohodlné, zavádíme *standardizované normální rozdělení* $N(0, 1)$, které získáme transformací hodnot veličiny x na hodnoty tzv. *standardizované náhodné veličiny* u vztahem

$$u = \frac{x - \mu}{\sigma}. \tag{1}$$

Hodnoty frekvenční a distribuční funkce rozdělení $N(0, 1)$ jsou pro praktické výpočty tabelovány, např. v [3].

Poněvadž naše empirické rozdělení četností náhodné veličiny můžeme pokládat pouze za určitou aproximaci normální distribuce (normální rozdělení vykazují totiž pouze spojité proměnné x, a to při nekonečně velkém počtu měření), nahradíme teoretické parametry μ, σ^2 odhady vypočtenými z našeho souboru. Podle [4] nejlepším odhadem parametru μ je *výběrový průměr* $\bar{x}$, který počítáme jako aritmetický průměr

$$\bar{x} = \frac{1}{n} \sum_1^k n_x x \tag{2}$$

a nejlepším odhadem parametru σ^2 *výběrový rozptyl*

$$s^2 = \frac{1}{n - 1} \sum_1^k n_x (x - \bar{x})^2 ; \tag{3}$$

standardní odchylka s je pak druhá odmocnina výběrového rozptylu.

Výběrové charakteristiky $\bar{x}$ a s^2 použijeme k testování statistických hypotéz. K testování hypotézy, že rozdělení četností náhodné veličiny X můžeme pokládat za normální, používáme *testu dobré shody.* K tomuto účelu rozdělujeme sledované hodnoty x do m tříd a absolutní četnost v i-té třídě označíme n_i. Smysl testu spočívá v tom, že hodnotíme pro jednotlivé třídy rozdíly mezi skutečnými četnostmi n_i a teoretickými četnostmi np_i, přičemž veličina p_i je pravděpodobnost, že četnost standardizované veličiny u_i definované vztahem (1) nabude očekávané hodnoty; při výpočtu hodnot u_i nahrazujeme parametry μ a σ^2 jejich odhady $\bar{x}$ a s^2. Testovacím kritériem je charakteristika χ^2, definovaná vztahem

$$\chi^2 = \sum_1^m \frac{(n_i - np_i)^2}{np_i}, \tag{4}$$

kde m je počet tříd. Hodnotu vypočtené charakteristiky porovnáváme pak s kritickou hodnotou $\chi_\alpha (f)$, kterou určíme pro daný počet stupňů volnosti f a pro zvolenou hladinu významnosti α z tabulek [3]. Počet stupňů volnosti je v tomto případě $f = m - 3$, tedy o 3 menší než počet tříd m, poněvadž u předpokládaného normálního rozdělení existují celkem tři omezení (součet teoretických četností je roven součtu skutečných četností, parametry μ a σ jsou ztotožněny s odpovídajícími výběrovými charakteristikami $\bar{x}$ a s). Pro hladinu významnosti α, která představuje určitý stupeň důvěry k rozhodnutí o přijetí či zamítnutí testované hypotézy, jsme předem zvolili hodnotu $\alpha = 0{,}05$. Hypotézu o normálním rozdělení pak nezamítáme, je-li $\chi^2 < \chi^2_\alpha(f)$.

Pro přesnější popis skutečného rozdělení používáme také někdy *koeficientu šikmosti* A (asymetrie) a *koeficientu špičatosti* E (exces), pro které se uvádějí [5] vztahy

$$A = \frac{\sum_1^k n_x(x - \bar{x})^3}{ns^3} \tag{5}$$

$$E = \frac{\sum_1^k n_x(x - \bar{x})^4}{ns^4} . \tag{6}$$

Pro normální rozdělení mají tyto koeficienty hodnoty $A = 0$, $E = 3$.

Charakteristiky s se také používá ke stanovení *spolehlivosti* (*reliability*) testů vědomostí. Na spolehlivost testu usuzujeme [6] z koeficientu

$$R = \frac{k}{k - 1} \frac{s^2 - \sum_1^k p_i q_i}{s^2} , \tag{7}$$

kde k je počet otázek testu, p_i jsou relativní četnosti správných odpovědí na jednotlivé otázky, q_i relativní četnosti nesprávných odpovědí. Spolehlivost testu roste s rostoucím koeficientem R, jehož maximální hodnota je rovna jedné.

Za předpokladu platnosti hypotézy o normálním rozdělení četností náhodné veličiny X a za předpokladu spolehlivosti použitého testu lze konstruovat *klasifikační normu* odpovídající výsledkům testu. Pro tuto normu volíme jednak klasifikační stupnici (v našem případě jsme použili čtyřstupňové klasifikace, která je obvyklá na vysokých školách), jednak procentové ekvivalenty odpovídající klasifikačním stupňům (byly zvoleny ekvivalenty 85%, 50%, 15%). K procentovým ekvivalentům přiřadíme meze náhodné veličiny X, tzv. *kvantily* x_p náhodné veličiny s rozdělením $N(\mu, \sigma^2)$, které počítáme vzhledem k (1) pomocí vztahu

$$x_p = \bar{x} + u_p s . \tag{8}$$

Hodnoty u_p jsou kvantily rozdělení $N(0, 1)$ a určíme je z tabulek.

Vzájemné srovnání výsledků dosažených chlapeckou a dívčí částí skupiny je provedeno *testem významnosti rozdílu mezi dvěma průměry*. Absolutní četnosti náhodné veličiny X označíme pro chlapce n_{x1}, pro dívky n_{x2}; celkový počet chlapců je n_1, celkový počet dívek n_2. Podle vztahů (2) a (3) určíme hodnoty výběrových průměrů $\bar{x}_1$ a $\bar{x}_2$ a příslušných variancí s_1^2 a s_2^2. O statistické významnosti rozdílu $\bar{x}_1 - \bar{x}_2$ rozhodneme pomocí *t-testu*; veličina t má tzv. Studentovo normální rozdělení s $f = n_1 + n_2 - 2$ stupni volnosti.

Výpočet testovacího kritéria t závisí však na tom, zda jsou či nejsou variance s_1^2 a s_2^2 obou výběrů homogenní. O této homogenitě variancí rozhodneme z poměru

$$F = \frac{s_1^2}{s_2^2}, \tag{9}$$

přičemž do čitatele dosazujeme vždy větší hodnotu rozptylu. Veličina F má tzv. Fisherovo-Snedecorovo rozdělení s $f_1 = n_1 - 1$ a $f_2 = n_2 - 1$ stupni volnosti a její kritické hodnoty jsou rovněž tabelovány. Hypotézu o homogenitě variancí zamítáme, je-li poměr F větší než kritická hodnota. V opačném případě hypotézu nezamítáme (tak tomu bylo právě v našem případě) a testovací kritérium t počítáme ze vztahu

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left[\dfrac{\sum_1^k n_{x1}(x_1 - \bar{x}_1)^2 + \sum_1^k n_{x2}(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}\right]}\sqrt{\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}. \tag{10}$$

Je-li vypočtená hodnota $|t| < t_\alpha(f)$, přičemž $t_\alpha(f)$ je kritická hodnota Studentovy distribuce, kterou vyhledáme pro zvolenou hladinu významnosti α opět v tabulkách [3], považujeme rozdíl v průměrech $\bar{x}_1$ a $\bar{x}_2$ za statisticky nevýznamný, v opačném případě za statisticky významný.

Korelace mezi výsledky prvního a druhého testu písemné zkoušky je vyjádřena *součinovým koeficientem korelace.* Hodnoty náhodné veličiny X označíme u prvého testu y, u druhého testu z a u každého posluchače sledujeme pak současně obě tyto hodnoty. Součinový koeficient korelace, který vyjadřuje stupeň těsnosti vztahu mezi proměnnými y a z, určíme podle [7] ze vztahu

$$r_{yz} = \frac{n \sum_1^n y_i z_i - \sum_1^n y_i \sum_1^n z_i}{\sqrt{[n \sum_1^n y_i^2 - (\sum_1^n y_i)^2][n \sum_1^n z_i^2 - (\sum_1^n z_i)^2]}}, \tag{11}$$

kde n je počet účastníků testu, y_i a z_i jsou dvojice hodnot, kvantifikující výsledky prvého a druhého testu u i-tého účastníka testu. Koeficient r může nabývat hodnot od 0 do ± 1; znaménko vyjadřuje kladný nebo záporný smysl vztahu mezi oběma proměnnými, absolutní hodnota korelačního koeficientu pak stupeň těsnosti tohoto vztahu. Podle [8] se hovoří obvykle o nepatrné korelaci, má-li $|r|$ hodnoty mezi 0 a 0,3, o střední korelaci pro $|r|$ mezi 0,4 a 0,6, o vysoké pro $|r|$ mezi 0,7 a 0,8 a o velmi vysoké korelaci pro $|r|$ nad 0,9. Při nepříliš vysokých hodnotách koeficientu korelace vzniká otázka, zda je korelace mezi zkoumanými veličinami opravdu významná, tj. neliší-li se hodnota vypočteného koeficientu r od nuly jen náhodně. O tom rozhodneme na základě porovnání hodnoty koeficientu korelace s kritickou hodnotou nalezenou pro zvolenou hladinu významnosti v příslušných tabulkách.

VÝSLEDKY PRVNÍHO TESTU

V prvním testu měli posluchači prokázat znalost jednotek deseti daných fyzikálních veličin (síla, tíha, práce, výkon, energie, hustota, tlak, frekvence, moment síly, moment setrvačnosti); pro každou veličinu určovali jednotku v soustavě SI, fyzikální

Tabulka I

x	n_x	n_{x1}	n_{x2}	n_i	n_{i1}	n_{i2}
3	5	1	4			
4	3	2	1	12	4	8
5	4	1	3			
6	2	2	—			
7	3	2	1			
8	8	4	4	22	11	11
9	3	2	1			
10	6	1	5			
11	5	4	1			
12	10	8	2			
13	8	2	6	42	21	21
14	11	3	8			
15	8	4	4			
16	4	1	3			
17	4	1	3			
18	8	4	4	26	12	14
19	5	4	1			
20	5	3	2			
21	6	4	2			
22	2	1	1			
23	3	2	1	18	12	6
24	1	1	—			
25	6	4	2			
26	—	—	—			
27	3	2	1			
28	—	—	—	7	6	1
29	1	1	—			
30	3	3	—			
Σ	127	66	61	127	66	61

rozměr jednotky a vztah, kterým je veličina definována a podle něhož lze odvodit rozměr jednotky. Test byl zadán na předtištěných formulářích a k jeho vypracování byla vymezena doba 10 minut. Správná odpověď na každou z 30 otázek testu byla hodnocena jedním bodem, každý posluchač mohl tedy získat maximálně 30 bodů.

Výsledky testu jsou uvedeny v tabulce I, kde v prvním sloupci jsou hodnoty x vyjadřující bodové výsledky testu, ve druhém jejich absolutní četnosti n_x, tj. počty posluchačů, kteří získali právě x bodů, ve třetím a čtvrtém sloupci četnosti n_{x1} pro chlapeckou část a n_{x2} pro dívčí část. Poslední tři sloupce tabulky obsahují absolutní četnosti ve třídách po pěti bodech, přičemž n_i je celková četnost v dané třídě, n_{i1} a n_{i2} četnosti v dané třídě pro chlapeckou a dívčí část.

Z číselných hodnot prvých čtyř sloupců byly podle vztahů (2) a (3) postupně vypočteny výběrové charakteristiky:

pro celou skupinu	pro chlapce	pro dívky
$\bar{x} = 14{,}71$	$\bar{x}_1 = 15{,}76$	$\bar{x}_2 = 13{,}57$
$s^2 = 43{,}45$	$s_1^2 = 50{,}68$	$s_2^2 = 33{,}85$
$s = 6{,}59$	$s_1 = 7{,}12$	$s_2 = 5{,}82$

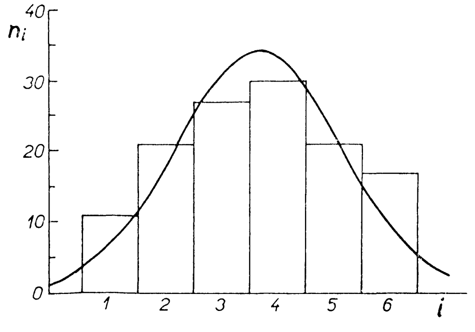
Výběrových charakteristik pro celou skupinu $\bar{x}$ a s a četností n_i v pátém sloupci tabulky I bylo použito k testu dobré shody. Nejdříve byly stanoveny podle vztahu (1) hodnoty standardizované veličiny u_i pro příslušné hranice tříd (tj. pro hodnoty $x = 5{,}5$; $10{,}5$ atd.), potom pomocí tabulek distribuční funkce normálního rozdělení

Tabulka II

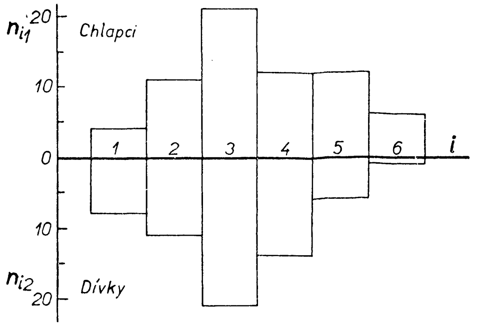
i	x	n_i	u_i	p_i	np_i	$n_i - np_i$
1	0— 5	12		0,0813	10,33	1,67
			−1,398			
2	6—10	22		0,1801	22,87	−0,87
			−0,639			
3	11—15	42		0,2864	36,37	5,63
			+0,120			
4	16—20	26		0,2625	33,34	−7,34
			0,879			
5	21—25	18		0,1389	17,64	0,36
			1,637			
6	26—30	7		0,0508	6,45	0,55
Celkem		127		1,0000	127,00	0,00

$N(0,1)$ pravděpodobnosti p_i a nakonec rozdíly mezi skutečnými četnostmi n_i a četnostmi teoretickými np_i, které by na danou třídu připadaly při normálním rozdělení. Přehled uvedených údajů je patrný z tabulky II.

K výpočtu testovacího kritéria χ^2 posloužily číselné hodnoty posledních dvou sloupců tabulky II. Jejich dosazení do vztahu (4) vedlo k výsledku $\chi^2 = 2{,}84$. Poněvadž tato vypočítaná hodnota je menší nežli kritická hodnota $\chi^2_\alpha(f) = 7{,}81$, určená pro $f = m - 3 = 3$ stupně volnosti na hladině významnosti $\alpha = 0{,}05$ z tabu-



Obr. 1. Histogram četností ve třídních intervalech u prvního testu pro celou skupinu.



Obr. 2. Histogram četností ve třídních intervalech u prvního testu pro chlapce a dívky.

lek [3], hypotéza o normálním rozdělení nebyla zamítnuta. Oprávněnost této hypotézy vyplývá také z histogramu na obr. 1. Na jeho vodorovnou osu jsou vyneseny třídy i náhodné veličiny X, na svislou osu četnosti n_i v jednotlivých třídách. Pro srovnání skutečného rozdělení s rozdělením normálním je histogram proložen křivkou frekvenční funkce odpovídající normálnímu rozdělení se stejnými charakteristikami $\bar{x}$ a s.

K dokreslení vlastností studovaného rozdělení byly určeny koeficient šikmosti A a koeficient špičatosti E. Výpočet koeficientu šikmosti podle (5) poskytl hodnotu $A = 0{,}27$, což znamená, že vrchol skutečného rozdělení je vzhledem k vrcholu normálního rozdělení, pro které je $A = 0$, posunut poněkud ve směru nižších hodnot x. Koeficient špičatosti podle vztahu (6) byl $E = 2{,}39$, tedy o něco menší nežli u normálního rozdělení, pro něž je $E = 3$.

Spolehlivost testu byla stanovena pomocí koeficientu R vypočteného podle vztahu (7), v němž jsme výraz $\sum_1^k p_i q_i$ nahradili výrazem $\sum_1^k P_i Q_i / n^2$, přičemž P_i jsou absolutní četnosti správných odpovědí na jednotlivé otázky testu, $Q_i = n - P_i$ absolutní četnosti nesprávných odpovědí. Četnosti správných odpovědí uvádíme v tabulce III. Po výpočtu $\sum_1^k p_i q_i = 5{,}52$ a po dosazení $k = 30$, $s^2 = 43{,}45$ dostáváme koeficient $R = 0{,}90$. Poněvadž je tato hodnota dosti vysoká, můžeme považovat náš test za dostatečně spolehlivý.

Tabulka III

	Síla	Tíha	Práce	Výkon	Energie	Hustota	Tlak	Frekvence	Moment síly	Moment setrvačnosti
Jednotka	118	26	101	94	72	52	24	76	17	7
Rozměr	85	43	60	43	51	74	29	51	28	9
Vztah	109	97	114	101	71	82	84	80	56	14

Na základě předchozích výsledků jsme oprávněni přistoupit k vytvoření klasifikační normy. Kvantily u_p standardizované náhodné veličiny korespondující se zvolenými procentovými ekvivalenty 85%, 50% a 15% jsou

$$u_{0,85} = 1{,}036\,, \quad u_{0,50} = 0{,}000\,, \quad u_{0,15} = -1{,}036\,,$$

a jim odpovídající kvantily x_p vypočtené podle vztahu (8)

$$x_{0,85} = 21{,}54\,, \quad x_{0,50} = 14{,}71\,, \quad x_{0,15} = 7{,}88\,.$$

Tím jsou také stanoveny meze jednotlivých klasifikačních stupňů: prvnímu klasifikačnímu stupni (výborně) odpovídá bodový zisk 30 – 22 bodů, druhému (velmi dobře) 21 – 15 bodů, třetímu (dobře) 14 – 8 bodů a čtvrtému (nevyhověl) 7 – 0 bodů. Vidíme, že body jsou na klasifikační stupně rozděleny téměř rovnoměrně; na první stupeň připadá rozmezí 9 bodů, na ostatní po 7 bodech. Nutno připomenout, že tato klasifikační norma, která se může jevit málo přísná, odpovídá jednak méně než průměrnému výsledku zkoušky ($\bar{x} = 14{,}71$), jednak předem zvolenému procentuálnímu rozdělení (procentovým ekvivalentům). Proto byli klasifikováni známkou velmi dobře ještě i ti studenti, kteří odpověděli správně jen na polovinu daných otázek.

Nyní provedeme srovnání výsledků dosažených odděleně chlapeckou a dívčí částí skupiny. Z tabulky I jsme vypočetli aritmetický průměr pro chlapce $\bar{x}_1 = 15{,}76$, pro dívky $\bar{x}_2 = 13{,}57$ a příslušné variance $s_1^2 = 50{,}68$, $s_2^2 = 33{,}85$. O homogenitě variancí se přesvědčíme dosazením hodnot s_1^2 a s_2^2 do vztahu (9). Takto vypočtenou hodnotu $F = 1{,}44$ porovnáme s kritickou hodnotou 1,53 pro $f_1 = 65$ a $f_2 = 60$ stupňů volnosti na pětiprocentní hladině významnosti a konstatujeme, že rozdíl ve variancích není ještě statisticky významný. Nyní testujeme významnost rozdílu v aritmetických průměrech, tj. významnost rozdílu $\bar{x}_1 - \bar{x}_2 = 2{,}19$. Testovací kritérium podle (10) je $t = 1{,}88$, kritická hodnota pro $f = n - 2 = 125$ stupňů volnosti je $t_\alpha(f) = 1{,}98$. Poněvadž je $|t| < t_\alpha(f)$, považujeme zjištěný rozdíl ve výsledcích testu dosažených chlapci a dívkami pouze za náhodný. Na obr. 2. je znázorněno rozdělení četností v jednotlivých intervalech pro chlapce a dívky (číselné hodnoty četností jsou v posledních dvou sloupcích tabulky I).

Vypočítáme-li ještě pro oba výběry koeficienty šikmosti $A_1 = 0{,}23$, $A_2 = 0{,}03$

a koeficienty špičatosti $E_1 = 2{,}34$, $E_2 = 2{,}51$, pozorujeme, že rozdělení pro chlapce se odchyluje od rozdělení normálního více než u dívek. Tyto odchylky však rovněž nelze pokládat za významné.

VÝSLEDKY DRUHÉHO TESTU

Druhý test písemné zkoušky obsahoval celkem 24 otázek (příklady, jednoduché fyzikální problémy, čtení grafů) z mechaniky; maximální bodový zisk tedy činil

Tabulka IV

x	n_x	n_{x1}	n_{x2}	n_i	n_{i1}	n_{i2}
1	1	—	1	11	1	10
2	2	—	2			
3	4	—	4			
4	4	1	3			
5	2	1	1	21	9	12
6	4	2	2			
7	8	4	4			
8	7	2	5			
9	6	2	4	27	9	18
10	4	2	2			
11	9	4	5			
12	8	1	7			
13	8	4	4	30	20	10
14	9	6	3			
15	10	9	1			
16	3	1	2			
17	5	2	3	21	13	8
18	3	2	1			
19	4	4	—			
20	9	5	4			
21	5	2	3	17	14	3
22	4	4	—			
23	5	5	—			
24	3	3	—			
Σ	127	66	61	127	66	61

24 bodů na jednoho posluchače. K vypracování testu byla vymezena doba 30 minut. Přehled o výsledcích poskytují tabulky IV a V, jejichž úprava je obdobná provedení tabulek I a II.

Tabulka V

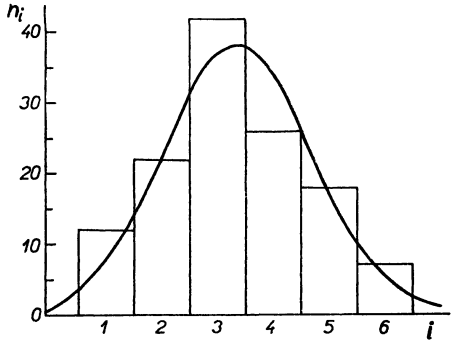
i	x	n_i	u_i	p_i	np_i	$n_i - np_i$
1	0– 4	11		0,0710	9,02	1,98
			−1,468			
2	5– 8	21		0,1437	18,25	2,75
			−0,790			
3	9–12	27		0,2406	30,56	−3,56
			−0,112			
4	13–16	30		0,2589	32,88	−2,88
			+0,566			
5	17–20	21		0,1790	22,73	−1,73
			+1,244			
6	21–24	17		0,1068	13,56	3,44
Celkem		127		1,0000	127,00	0,00

Výpočet výběrových charakteristik z číselných podkladů tabulky IV dává tyto hodnoty:

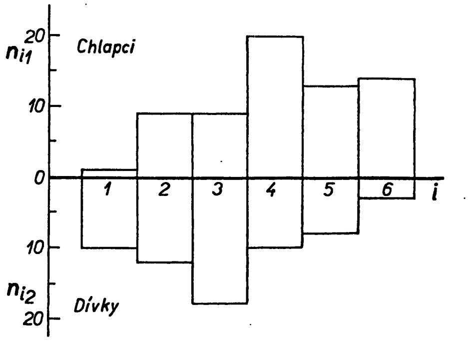
pro celou skupinu	pro chlapce	pro dívky
$\bar{x} = 13{,}16$	$\bar{x}_1 = 15{,}29$	$\bar{x}_2 = 10{,}85$
$s^2 = 34{,}87$	$s_1^2 = 30{,}69$	$s_2^2 = 29{,}57$
$s = 5{,}90$	$s_1 = 5{,}54$	$s_2 = 5{,}43$

Potřebné údaje k realizaci testu dobré shody obsahuje tabulka V. Vypočítaná hodnota testovacího kritéria $\chi^2 = 2{,}52$ je opět menší než kritická hodnota $\chi_\alpha^2(f) = 7{,}81$, proto ani zde nezamítáme hypotézu o normálním rozdělení. Navíc z rozdílů mezi skutečnými a teoretickými četnostmi třídních intervalů, které lze sledovat v posledním sloupci tabulky, je zřejmé, že skutečné rozdělení je poněkud plošší než normální (v okrajových třídách jsou rozdíly kladné, ve středních záporné). O tom svědčí také hodnota koeficientu špičatosti $E = 2{,}09$, která je menší než u prvého testu. Koeficient šikmosti $A = 0{,}016$ je rovněž menší než u prvého testu; jeho hodnota je velmi blízká hodnotě koeficientu normálního rozdělení. Uvedené skutečnosti dokumentuje názorně histogram skutečného rozdělení na obr. 3, který je opět proložen Gaussovou křivkou.

V tabulce VI jsou uvedeny absolutní četnosti správných odpovědí na jednotlivé otázky testu, kterých použijeme opět ke stanovení spolehlivosti testu. Poněvadž výpočet koeficientu podle (7) dává hodnotu $R = 0{,}89$, je spolehlivost testu prakticky stejná jako spolehlivost testu prvého; tedy také druhý test můžeme pokládat za dostatečně spolehlivý.



Obr. 3. Histogram četností ve třídních intervalech u druhého testu pro celou skupinu.



Obr. 4. Histogram četností ve třídních intervalech u druhého testu pro chlapce a dívky.

Procentové ekvivalenty pro stanovení kvantilů ke klasifikaci výsledků testu jsme volili stejné jako u prvního testu. Kvantily u_p standardizované veličiny zůstávají

$$u_{0,85} = 1{,}036\,, \quad u_{0,50} = 0{,}000\,, \quad u_{0,15} = -1{,}036\,,$$

jim odpovídající kvantily x_p jsou

$$x_{0,85} = 19{,}27\,, \quad x_{0,50} = 13{,}16\,, \quad x_{0,15} = 7{,}05\,.$$

Tabulka VI

Pořadové číslo otázky	1	2	3	4	5	6	7	8	9	10	11	12
Počet správných odpovědí	118	95	93	91	82	105	52	93	77	77	59	85
Pořadové číslo otázky	13	14	15	16	17	18	19	20	21	22	23	24
Počet správných odpovědí	40	35	64	45	66	71	92	37	54	56	53	31

Z hlediska rozmístění kvantilů na ose x jsou výsledky druhého testu poněkud lepší. Prvému klasifikačnímu stupni odpovídá bodový zisk 24—20 bodů, druhému 19—14 bodů, třetímu 13—8 bodů a čtvrtému 7—0 bodů; posluchač, který odpověděl správně na polovinu otázek testu, je zde klasifikován známkou dobře.

Rozdíl v aritmetických průměrech dosažených chlapci a dívkami je tentokráte podstatně větší než u prvního testu. Vzhledem k vypočteným hodnotám výběrových charakteristik $\bar{x}_1 = 15{,}29$, $\bar{x}_2 = 10{,}85$ a $s_1^2 = 30{,}69$, $s_2^2 = 29{,}57$ je rozdíl průměrů $\bar{x}_1 - \bar{x}_2 = 4{,}44$, což při značné homogenitě obou variancí $(F = 1{,}04)$ dává hodnotu testovacího kritéria $t = 5{,}31$. Jelikož kritická hodnota na pětiprocentní hladině významnosti činí jen $t_\alpha(f) = 1{,}98$, jsme oprávněni konstatovat, že zjištěný rozdíl mezi průměry je statisticky významný. Značného rozdílu ve výkonu chlapců a dívek jsme si mohli také povšimnout u četností n_{i1} a n_{i2} v tabulce IV. Ještě výrazněji se tato skutečnost projevuje na histogramu četností, znázorněném na obr. 4.

Koeficient špičatosti E je pro obě dílčí rozdělení prakticky stejný a blízký hodnotě 2. Koeficienty šikmosti jsou co do absolutní hodnoty rovněž skoro stejné, liší se však znaménkem (pro chlapce $A_1 = -0{,}16$, pro dívky $A_2 = 0{,}17$), což znamená, že vrchol rozdělení u chlapců je posunut směrem k vyšším hodnotám x, u dívek směrem k nižším hodnotám.

KORELACE MEZI VÝSLEDKY PRVNÍHO A DRUHÉHO TESTU

Vzhledem k vysoce významnému rozdílu ve výsledcích druhého testu u chlapců a dívek jsme určovali součinový koeficient korelace pro obě skupiny zvlášť. Ke každému z n_1 chlapců a z n_2 dívek jsme přiřadili dvojici hodnot y_i (výsledek prvního testu) a z_i (výsledek druhého testu) a pomocí vztahu (11) vypočetli hodnoty součinového koeficientu korelace: pro chlapce $r_1 = 0{,}67$, pro dívky $r_2 = 0{,}48$. Obě hodnoty jsou významné na pětiprocentní hladině významnosti; kritická hodnota pro $n_1 = 66$ je 0,24, pro $n_2 = 61$ pak 0,25. Z vyšší hodnoty koeficientu r_1 lze usoudit, že těsnější vztah mezi výsledky prvého a druhého testu je u chlapecké skupiny.

ZÁVĚR

Z uvedených výsledků písemné kontrolní zkoušky z fyziky, provedené u 127 absolventů střední školy, lze celkově usuzovat na několik závažných skutečností.

1. Projevil se relativně nízký rozsah i kvalita vědomostí absolventů v základních partiích fyzikálního učiva. Průměrný výsledek prvého testu činil 14,71 bodů ze 30 bodů možných, což znamená, že průměrný posluchač neodpověděl správně ani na polovinu otázek testu; výsledek druhého testu byl 13,16 bodů ze 24. K tomu pro zajímavost uvádíme, že z absolventů SVVŠ bylo klasifikováno z fyziky na závěrečném vysvědčení 69 studentů výborně, 40 chvalitebně a jen 8 dobře (10 studentů přišlo

z průmyslových škol, kde se fyzika ve vyšších ročnících nevyučuje), přičemž z fyziky maturovalo celkem 72 studentů.

2. Ve vědomostech chlapecké a dívčí skupiny se projevily podstatné rozdíly. Rozdíl v průměrných výsledcích chlapců a dívek u prvého testu byl 2,19 bodu (na zvolené pětiprocentní hladině významnosti jsme jej však nepokládali za signifikantní), rozdíl v průměrných výsledcích druhého testu 4,44 bodu, což činí 18,5% variační šíře 24 bodů; statistická významnost tohoto rozdílu byla objektivně prokázána. Vzhledem k zaměření druhého testu lze říci, že v neformálních znalostech chlapci výrazně předčili dívky. Tento závěr je zcela neočekávaný, neboť průměrná známka z fyziky na výročním vysvědčení byla u chlapců 1,62 a u dívek 1,40; vzhledem k tomu bychom očekávali lepší výsledky spíše u dívek.

3. Mezi výsledky prvého a druhého testu byla zjištěna pouze střední korelace, přičemž těsnější vztah mezi oběma testy vykazovala chlapecká skupina.

Skutečnosti uvedené ad 2 by zasloužily hlubšího průzkumu, zvláště s ohledem na zřejmý nesoulad výsledků zjištěných u druhého testu s hodnocením vědomostí chlapců a dívek ve fyzice na střední škole. Toto zjištění je zajímavé i vzhledem k tomu, že ve studiu na vysoké škole se významné rozdíly mezi chlapci a dívkami dále zřetelně neprojevují. V dalších testech, jejichž obsahem bylo již vysokoškolské učivo fyziky, jsme totiž mohli zjistit postupné vyrovnávání výkonů u obou skupin.

Literatura

[1] Bednařík M. - Široká M.: Výsledek výzkumu vědomostí z fyziky u skupiny absolventů střední školy. Fyzika ve škole *8*, 1969, č. 2, s. 90—98.

[2] Komenda S.: *Základy pravděpodobnostních a statistických metod v psychologickém a pedagogickém výzkumu.* Učební texty vysokých škol, SPN, Praha 1967.

[3] Janko J.: *Statistické tabulky.* ČSAV, Praha 1958.

[4] Lindquist E. F.: *Statistická analýza v pedagogickém výzkumu.* SPN, Praha 1967.

[5] Itelson L.: *Mathematische und kybernetische Methoden in der Pädagogik.* Volk und Wissen Volkseigener Verlag, Berlin 1967.

[6] Guilford J. P.: *Podstawowe metody statystyczne w psychologii i pedagogice.* Państwowe wydawnictwo naukowe, Warszawa 1964.

[7] Mittenecker E.: *Plánování a statistické hodnocení experimentů.* SPN, Praha 1968.

[8] Meili R. - Rohracher H.: *Učebnice experimentální psychologie.* SPN, Praha 1967.