



ПРИЛОЖЕНИЕ НА МЕТОДА НА ШЕФЕ В ОБРАЗОВАНИЕТО И ТРАНСПОРТА

Драго Михалев, Райна Алашка
michalev@abv.bg, alraina@abv.bg

*ВТУ „Тодор Каблешков”, София 1574, ул. „Гео Милев”158,
БЪЛГАРИЯ*

Ключови думи: образование, транспорт, статистика, метод на Шефе, дисперсионен анализ, приложна математика.

Резюме: Дадени са основните термини, формули и стъпки на дисперсионния анализ. Дисперсионният анализ е метод за изследване на различни зависимости. Описан е методът на Шефе. Той е по-добър от метода на дисперсионния анализ. С него могат да се направят сравнения на всеки две групи или да се направят сравнения на комбинация от няколко групи. При двата метода се използва F -разпределение на Фишер. Приложен е методът на Шефе за резултатите на студентите в зависимост от формата на обучението им. Направено е сравнително изследване относно ефективността на четири различни форми на обучение. Статистическите данни са получени от резултатите на 79 студента от ВТУ“Тодор Каблешков“. Показана е групата, която дава най-добър резултат. Дадени са примери за възможно приложение на метода в транспорта и образованието.

1. ДИСПЕРСИОНЕН АНАЛИЗ

Дисперсионният анализ е метод за проверка на хипотеза за равенство на повече от две средни едновременно [1], [2], [3]. Дисперсионният анализ е метод за изследване на зависимости. Чрез него може да се провери има или не зависимост, но не може да се прецени колко силна е тази зависимост. При прилагането на дисперсионният анализ се използват различни компоненти на дисперсията, но се прави извод за равенство или не на средните, получени при различни значения на фактора–причина.

Нека изследвания фактор–причина X приема k различни значения, които определят броя k на групите. Броят на елементите в j -тата група ще отбелязваме с n_j .

Основните понятия и означения в дисперсионния анализ са:

Общ брой на всички елементи: $N = \sum_{j=1}^k n_j$. Средно в j -тата група: $\bar{y}_j = \frac{\sum_{i=1}^{n_j} y_{ij}}{n_j}$.

Общо средно: $\bar{y} = \frac{\sum_{j=1}^k \bar{y}_j}{k}$. Обща девиация: $SS_o = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2$.

Вътрешногрупова девиация: $SS_b = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$. Междугрупова девиация:

$$SS_m = \sum_{j=1}^k n_j (\bar{y}_j - \bar{y})^2.$$

При дадено ниво на значимост α , теоретичната стойност F_T се взема от таблицата за F -разпределение. В сила е равенството $F_T = F(\alpha, k-1, N-k)$.

Един удобен начин за представяне на получените резултати е стандартната таблична форма (Таблица 1). Този начин е стандартен за повечето статистически пакети.

Таблица 1

Източник на дивияция	Сума от квадратите на отклоненията	Степени на свобода	Оценки на дисперсията	Емпирична стойност	Теоретична стойност
Междугрупова	SS_m	$k-1$	$s_m^2 = \frac{SS_m}{k-1}$	$F_e = \frac{s_m^2}{s_b^2}$	$F_T = F(\alpha, k-1, N-k)$
Вътрешногрупова	SS_b	$N-k$	$s_b^2 = \frac{SS_b}{N-k}$	$F_e \leq F_T \Rightarrow$ Приемаме H_0 , $F_e > F_T \Rightarrow$ Приемаме H_1 .	
Обща	SS_o	$N-1$	$s_o^2 = \frac{SS_o}{N-1}$		

2. МЕТОД НА ШЕФЕ

Методът на Шефе (S-методът) се препоръчва, когато чрез дисперсионния анализ е установено наличието на значимо F -отношение. S-методът има предимството, че могат да се правят не само подвойкови сравнения, а и при сложна структура на необходимите сравнения.

Нека имаме k групи и искаме да ги сравним при ниво на значимост α . S-методът минава през два етапа:

1. Проверява се с дисперсионния анализ нулевата хипотеза $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ при даденото ниво на значимост α , ако нулевата хипотезата се отхвърли се преминава на втория етап;
2. Създава се необходимият контраст. Изчислява се по формулата емпиричната стойност \bar{F}_e и теоретичната стойност \bar{F}_T и се сравняват.

При метода на Шефе се създават контрасти, които са линейна комбинация на средните с коефициенти, чиято сума е нула. Видът на един контраст е:

$$(1) \quad \sum_{i=1}^k C_i \mu_i, \text{ където } \sum_{i=1}^k C_i = 0.$$

В зависимост от това какво искаме да проверим контрастите могат да имат различен вид, но винаги трябва сумата от коефициентите да е нула. Приравняват се контрастите на нула и получаваме вида на нулевите хипотези.

В обобщен вид нулевата хипотеза се записва като:

$$(2) \quad H_o: \sum_{i=1}^k C_i \mu_i = 0, \text{ където } \sum_{i=1}^k C_i = 0.$$

Когато някой от коефициентите $C_j = 0$, това означава, че j -тата група се игнорира от сравнението.

Тестовата статистика (емпиричната стойност) по метода на Шефе е:

$$(3) \quad \bar{F}_e = \frac{\left(\sum_{j=1}^k C_j \bar{y}_j \right)^2}{s_b^2 \left(\sum_{j=1}^k \frac{C_j^2}{n_j} \right)},$$

където s_b^2 е оценката на вътрешногруповата дисперсия от дисперсионния анализ, C_j са коефициентите на контраста и n_j е обема на j -тата група.

За теоретичната стойност имаме $\bar{F}_T = (k-1)F_T$, където F_T е теоретичната стойност от дисперсионния анализ (F -разпределението).

Когато $\bar{F}_e > \bar{F}_T$ нулевата хипотеза се отхвърля и се приема алтернативната хипотеза. Ако това неравенство не е изпълнено се приема нулевата хипотеза.

3. ОЦЕНКА НА РАЗЛИЧНИТЕ ФОРМИ НА ОБУЧЕНИЕ

Ще направим сравнително изследване относно ефективността на четири различни форми на обучение.

Разполагаме с резултатите от изпита по ВМ2 на студенти от ВТУ „Тодор Каблешков“ обучавани по четири различни форми на обучение: задочно обучение със задължителна домашна работа (форма 1), редовно обучение със задължителна домашна работа (форма 2), редовно обучение без задължителна домашна работа (форма 3) и задочно обучение без задължителна домашна работа (форма 4).

Цел на изследването е установяване наличието или отсъствието на зависимост между резултатите от изпита по ВМ 2 и формата на обучение.

Място на изследването е ВТУ „Тодор Каблешков“.

Обект на изследването са 79 студента – редовно и задочно обучение от I курс.

Предмет на изследването са резултатите от изпита по ВМ 2 част.

Данните са получени от изпитните работи и изпитните протоколи.

Хипотеза за резултата от изследването: Формата на обучение оказва влияние на резултатите от изпита по ВМ 2 част.

Методи на теоретичното изследване: Дисперсионен анализ, Метод на Шефе.

Получените резултати са представени в таблица 2.

Таблица 2

Форма на обучение	Брой точки от изпита по ВМ2														
	145	170	65	105	120	62	70	68	80	79	42	30	26	18	21
Задочно обучение -ДР	50	12	35	24	10	0									
Редовно обучение -ДР	135	142	153	98	60	65	75	80	64	32	40	25	42	36	0
Редовно обучение	150	99	100	63	75	72	69	43	27	39	48	45	0	0	
Задочно обучение	76	88	62	26	60	58	69	43	27	20	32	45	0	0	0
Задочно обучение	60	65	14	12	16	15	10	0	0	0	0	0	0		

Нека означим средния брой точки, получени от студентите, обучавани по първа, втора, трета и четвърта форма на обучение, съответно с $\mu_1, \mu_2, \mu_3, \mu_4$.

Дефинираме нулевата хипотеза $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$.

Алтернативната хипотеза е $H_1: \mu_1 \neq \mu_2$ или $\mu_1 \neq \mu_3$ или $\mu_1 \neq \mu_4$
 $\mu_2 \neq \mu_3$ или $\mu_2 \neq \mu_4$ или $\mu_3 \neq \mu_4$.

Средният брой точки на поне две от формите на обучение се различават значимо.
 Равнище на значимост $\alpha = 0,05$.

В таблица 3 са пресметнати основните характеристики по групи: общ брой единици в групата, сумата на всички измервания в групата, средното аритметично и дисперсията в групата.

Таблица 3

Форма	Група	Брой	Сума	Средно	Дисперсия
Форма 1	Задочно обучение със задължителна домашна работа	21	1232	58,67	2097,83
Форма 2	Редовно обучение със задължителна домашна работа	29	1877	64,72	1801,35
Форма 3	Редовно обучение	16	606	37,88	855,98
Форма 4	Задочно обучение	13	192	14,77	492,53

В таблица 4 са представени резултатите от еднофакторния дисперсионен анализ.

Таблица 4

Източник на диваация	Сума от квадратите на отклоненията	Степени на свобода	Оценки на дисперсията	Емпирична стойност	Теоретична стойност
Междугрупова	26329,1	3	8776,36	$F_e = 5,92$	$F(0,05, 3, 75) = F_T = 2,73$
Вътрешногрупова	111144,5	75	1481,93	$F_e > F_T \Rightarrow$ Приемаме H_1 .	
Обща	137473,6	78			

Тъй като $F_e > F_T$, отхвърляме нулевата хипотеза и приемаме алтернативната.

Изводът е, че ефектът от обучението по четирите различни форми е различен. Не е ясно обаче кои именно форми или групи от форми се различават и дали има форми, които са с еднакъв ефект.

За да дадем отговор на горните въпроси ще приложим метода на Шефе.

Имаме 4 форми на обучение и искаме да сравним всяка форма с всяка от останалите. Възможният брой на нулевите хипотези е $C_4^2 = \binom{4}{2} = 6$

След пресмятане по Метода на Шефе получаваме следните резултати за връзката между формите на обучение.

Таблица 5

Група	1	2	3
1			
2	H_0		
3	H_0	H_0	
4	H_1	H_1	H_0

Изводът е, че ефектът от обучението по форма 4 е различен от обучението по форма 1 и форма 2.

Сега ще проверим дали обучението по форма 2 (редовно обучение със задължителна домашна работа) дава резултати, които се различават от резултатите на другите три форми на обучение. Записваме нулевата хипотеза във вида:

$$H_0 : (+1)\mu_1 + (-3)\mu_2 + (+1)\mu_3 + (+1)\mu_4 = 0.$$

В този сложен контраст всички коефициенти са различни от нула и всички средни участват в изчисленията. В таблица 6 са дадени необходимите данни и изчисления за този контраст.

Таблица 6

Група (форма на обучение)	1	2	3	4
Средно в групата \bar{y}_j	58,67	64,72	37,88	14,77
Брой в групата n_j	21	29	16	13
Коефициент	+1	-3	+1	+1
Емпирична характеристика	$\bar{F}_e = \frac{(58,67 - 3 \cdot 64,72 + 37,88 + 14,77)^2}{1481,93 \cdot \left(\frac{1}{21} + \frac{9}{29} + \frac{1}{16} + \frac{1}{13} \right)} = 9,32$			

Сравняваме съответната емпирични характеристики с теоретичната. Тъй като $9,32 = \bar{F}_e > \bar{F}_T = 8,18$, нулевата хипотеза се отхвърля и се приема алтернативната хипотеза. Можем да направим *извод*, че обучението по форма 2 дава среден резултат, който е различен (по-добър) от средния резултат на комбинацията от другите три форми.

ЗАКЛЮЧЕНИЕ:

Възможни са и други приложения в образованието, като се отчита методиката на преподаване. Ще дадем и примери за приложение на метода на Шефе в транспорта.

Да разгледаме една автобусна фирма по един маршрут и оценяваме броя на пътниците според деня от седмицата или наблюдаваме продължителността и броя на пътниците на един курс по една и съща линия според часовата зона на деня и други.

Посочените примери водят до извода, че е необходимо студентите да изучават статистически методи, защото те ще им бъдат полезни в бъдещата професия.

ЛИТЕРАТУРА:

- [1] Михалев Д. Й., Алашка Р. М. Теория на вероятностите и статистика., София, 2012 г.
- [2] Калинов Кр. Статистически методи в поведенческите и социални науки., София, НБУ, 2010г.
- [3] Шеффе Д. Дисперсионный анализ., Москва, Фиматгиз, 1963г.

APPLICATION OF SCHEFFE'S METHOD IN EDUCATION AND TRANSPORT

Drago Michalev, Rayna Alashka
michalev@abv.bg, alraina@abv.bg

*Todor Kableshkov University of Transport,
158 Geo Milev Street, Sofia,
BULGARIA*

***Key words:** Education, transport, statistic, Scheffe's method, dispersion analysis, applied mathematic*

***Abstract:** The basic terms, formulas and steps of variance analysis are given. Variance analysis is method for research of different dependencies. Scheffe's method is presented. In the Scheffe's method, contrasts are created, which are a linear combination of the average with coefficients whose sum is zero. The method is better than method of variance analysis. With it we can to do congruence of every two groups and to do congruence of combination of some groups. In both methods F-distribution (Fisher's distribution) is applied. Scheffe's method for score of students depending of education's form is applied. A comparative study has been conducted on the effectiveness of four different forms of training. The statistical data are obtained from score of seventy nine students in Higher Transport School "Todor Kableshkov". The group, which give the best results is shown. Some examples for possible applications of the method in transports and education are given. For example, let's look at a bus company on one route and we estimate the number of passengers by day of the week, or we observe the length and number of passengers per course on the same line according to the time zone of the day and others.*