

МОЖЛИВОСТІ АНАЛІЗУ SPSS STATISTICS ДЛЯ ОПРАЦЮВАННЯ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ

Михайло Малезжик, Максим Привольнев

Національний педагогічний університет імені М. П. Драгоманова

Анотація:

Стаття містить опис застосування пакета статистичного аналізу даних для автоматизації розрахунку статистичних показників за класичною теорією. Проведено розрахунок основної сукупності статистичних показників (складність, дискримінативність, кореляція, надійність Альфа Кронбаха) для визначення якості тестового завдання. Наведено опис можливості застосування вбудованої у SPSS мови програмування, використано макроси й файли синтаксису (для розрахунку складності й дискримінативності). Отримані результати свідчать про ефективність застосування статистичного пакета SPSS для обробки результатів тестування.

Ключові слова:

аналіз; статистика; результати тестування; синтаксис; психометрія.

Анотация:

Малезжик Михаил, Привольнев Максим. **Возможности анализа SPSS STATISTICS для обработки результатов тестирования.** В статье содержится описание использования пакета статистического анализа данных для автоматизации расчета статистических показателей по классической теории. Произведен расчет основной совокупности статистических показателей (сложность, дискриминативность, корреляция, надежность Альфа Кронбаха) для определения качества тестового задания. Описаны возможности применения встроенного в SPSS языка программирования, использованы макросы и файлы синтаксиса (для расчета сложности и дискриминативности). Результаты свидетельствуют об эффективности применения статистического пакета SPSS для обработки результатов тестирования.

Ключевые слова:

анализ; статистика; результаты тестирования; синтаксис; психометрия.

Resume:

Malezhyk Mykhailo, Pryvolnev Maksym. **Possibilities of the analysis of SPSS STATISTICS for test results processing.**

This paper describes the use of statistical analysis package to automate the calculation of statistics by the classical theory. The calculation of the basic set of statistical indicators (P-value, D-index, correlation, reliability, Cronbach's alpha) for determining the quality of the test task has been made. The authors describe the possibility of using the built-in programming language in SPSS, using macros and syntax files (for P-value and D-index calculation) and thereby prove the effectiveness of the statistical package SPSS for test results processing.

Key words:

analysis; statistics; test results; syntax; psychometry.

Постановка проблеми. У міжнародній практиці психометрія стала невіддільним складником процесу розроблення тестів, оскільки її результати на етапі конструювання тесту дають змогу винести обґрунтоване рішення стосовно тестових завдань. Оскільки для одного тесту може бути створена велика кількість завдань, то розрахунки для завдань усього тесту супроводжуються великими витратами часу й похибками вимірювання, що виникають через неухважність або незнання понятійного апарату. Для уникнення цього використовують програмні засоби (SAS, SPSS Statistics, STATISTICA, R, Parscale тощо), проте для роботи з ними необхідно мати певний рівень підготовки.

Пакет SPSS Statistics зі статистичної обробки й аналізу даних стане в пригоді під час проведення практичних експериментів, які є невіддільною частиною роботи в галузі тестування.

Аналіз останніх досліджень і публікацій. Детальний аналіз даних за допомогою пакетів Statistica, SPSS Statistics та Excel розглянуто в навчальних посібниках В. Бахрушина [1], І. Лупан у співавторстві з О. Авраменко [3]. Досвід опрацювання психометричних даних у SPSS Statistics висвітлили такі дослідники, як А. Соколов [7], С. Раков [6], А. Лагунова [2] й М. Мазорчук [4].

Формулювання цілей статті. Зважаючи на стан розробки проблеми, метою статті є розкриття особливостей використання

аналітичних можливостей SPSS Statistics, необхідних під час розробки тестів.

Виклад основного матеріалу дослідження. SPSS Statistics не орієнтований на конкретний напрямок і має потужний функціонал, який застосовують до різноманітної кількості досліджень. Використовуючи цей статистичний пакет, користувач працює в одному з таких вікон, як: перегляд даних, перегляд змінних, перегляд вихідних даних, редактор синтаксису (застосовується для автоматизації), щоб зберегти або уточнити свої запити.

Статистичний пакет SPSS Statistics поєднує в собі функції введення й управління даними, роботу з графіками, таблицями, вивід інформації у формі web-сторінок і дає змогу працювати з даними, презентованими в різноманітних форматах, з мітками значень, а також робити перетворення даних, розщеплення й сортування. Статистичний апарат містить кілька сотень процедур і зручних інструментів виведення результатів – тривимірні гістограми й точкові діаграми (щільності точок або «хмари»). Є модуль , що дає змогу будувати дерева класифікацій, ідентифікувати групи й модулі для дисперсійного аналізу й дослідження регресійної моделі [5].

У цій статті розглянемо деякі можливості застосування пакета SPSS Statistics для опрацювання результатів тестування в педагогічних дослідженнях. Для визначення якості тестового завдання важливими є такі

статистичні дослідження, як нормальність розподілу, P-value (складність), D-index (дискримінативність), точково-бісеріальний коефіцієнт кореляції, надійність Альфа Кронбаха.

SPSS Statistics відзначається потужним статистичним функціоналом, однак не дуже зручними є внесення й редагування змінних, на відміну від MS Excel, тому дані рекомендується вносити останнім аналітичним пакетом. SPSS Statistics добре узгоджений з MS Excel, завдяки чому можна відкрити будь-який потрібний лист з книги формату MS Excel.

Визначення нормальності розподілу виконується за допомогою команди «Розвідковий аналіз», яку можна застосовувати до однієї чи двох груп одночасно, тобто можна

додавати незалежну змінну, що корисно, коли декілька класів пройшли тестування із зазначенням цього в окремій змінній. SPSS Statistics виводить багато таблиць і графіків з цією процедурою. Однією з причин цього є те, що команда «Розвідковий аналіз» застосовується не тільки для перевірки нормальності, а й для опису даних багатьма різними способами. Перевірка нормальності розподілу проводиться за двома добре відомими критеріями нормального розподілу Колмогорова-Смирнова й Шапіро-Вілка (табл. 1) і графіка розподілу нормальності (рис. 1). У ході дослідження було використано дані математичного складника тесту ТЗНК (тест загальної навчальної компетентності), пройденого учнями 11 класу в 2014 році в м. Києві.

Таблиця 1

Критерії нормальності, отримані на основі даних проведеної математичної частини тесту ТЗНК (вбірка з 26 учасників). Тест Шапіро-Вілка й Колмогорова-Смирнова

	Колмогоров-Смирнов ^а			Шапіро-Вілк		
	Статистика	ст. св.	Значущість	Статистика	ст. св.	Значущість
SUM	,120	26	,200*	,957	26	,344

а. Поправка значущості Ліллієфорса

*. Це нижня межа для істинної значущості

Тест Шапіро-Вілка найкраще застосовувати до малих розмірів вибірки (<50 тестованих), однак його можна використовувати й для великих розмірів вибірки (2000 тестованих). Як видно з таблиці, сумарні бали змінної SUM (сумарна кількість сирих балів) нормально розподілені. Якщо

«Значущість» тесту Шапіро-Вілка перевищує 0,05, то дані нормально розподілені. Якщо «Значущість» нижча за 0,05, то дані значно відхиляються від нормального розподілу. Однак на графіку нижче можна побачити, що більшість учнів набрала понад 25 сирих балів, тоді як 15 і 30 балів не набрав ніхто.

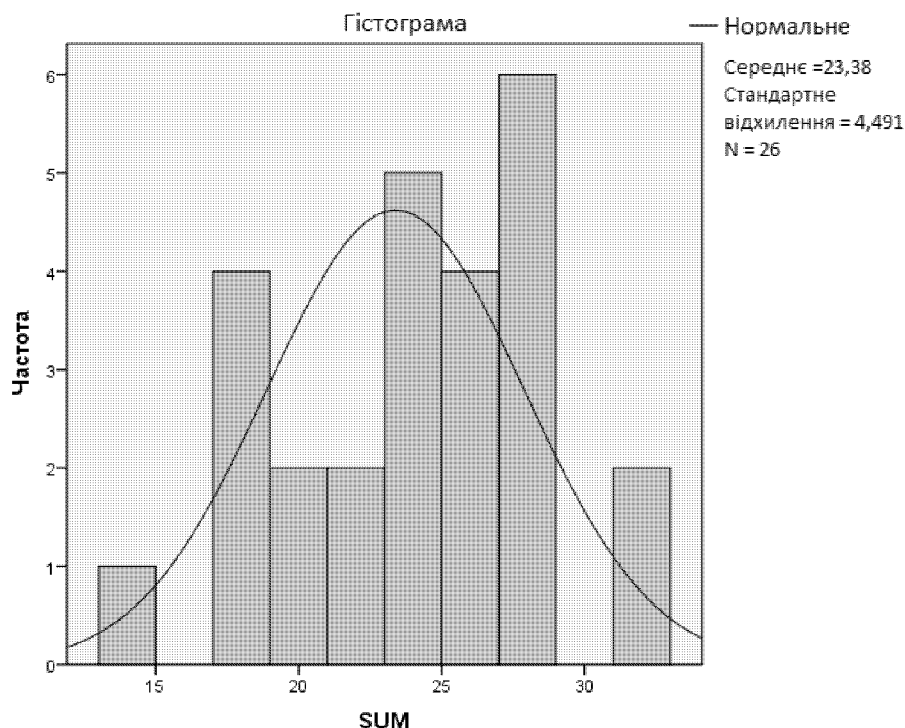


Рис. 1. Графік нормальності розподілу, отриманий на основі даних проведеної математичної частини тесту ТЗНК (вбірка з 26 учасників)

Складність і дискримінативність завдань тесту можна визначити за допомогою SPSS Statistics, проте для цього знадобиться використання синтаксису. Пуї-Ва Леї й Кіонг Ву створили макрос CTTITEM SAS і синтаксис SPSS, щоб обчислити складність і дискримінативність. Так, зокрема, аналіз застосовується не тільки до дихотомічних, а й до політомічних завдань. Крім того, D-index додатково обчислює дихотомічні дані завдань. Програми є зручними для користувача завдяки зменшенню кількості необхідних вхідних модифікацій і надання коротких навчальних посібників [8].

Ця розробка прискорює роботу над аналізом і виводить не тільки таблицю зі складністю завдань та їх дискримінативністю (рис. 2), а й опрацьовує відповіді учасників тестування, створюючи таблиці поділу дистракторів, що також є дуже корисним.

Користуючись розробкою Пуї-Ва Леї та Кіонг Ву необхідно керуватися їхніми рекомендаціями, що надані в навчальних посібниках, збережених в архівах з самими файлами для аналізу, а також у власній статті [8]. Вхідні дані повинні бути в текстовому форматі. Перший рядок має містити правильні відповіді на дихотомічні завдання або кількість категорій балів для політомічних завдань (програма передбачає, що ваші відповіді/оцінки для політомічних завдань починаються з 1 з інтервалом в один бал). Для дихотомічних завдань дані відповідей на завдання починаються з другого рядка й з кожною лінією презентують окреме завдання. Для політомічних завдань другий рядок повідомляє програмі, чи слід оцінювати завдання в зворотному порядку так (0) чи ні (1). Бали випробуваних починаються з третього рядка для політомічних

завдань. Використовуйте пробіли, якщо значення немає. Для дихотомічних завдань значення, якого немає, розглядатиметься як неправильне для розрахунку статистики завдань і тесту. Коли немає значення в політомічних завданнях, то видаляються учасники вибірки. Програма буде оцінювати дихотомічні елементи як правильні (надавши кодове значення «1») або неправильні (кодуються як «0») на основі ключів відповідей, наданих користувачем. До синтаксису повинні бути внесені деякі зміни, щоб вони відповідали поточній структурі даних. Ці зміни, пов'язані з розташуванням файлів та іменами змінних, також містяться в інструкціях у самому файлі синтаксису.

Перша частина виводу даних містить у собі статистику загального розподілу балів. Результати аналізу надійності, які йдуть за загальним розподілом балів, охоплюють надійність Альфа Кронбаха, зведену статистику з кореляції між завданнями й загальну статистику за завдання. «Кореляцію пункту з сумарним балом» зазвичай називають скоригованою точково-бісеріальною кореляцією. Потім перераховуються складність і дискримінативність для кожного завдання в його початковому порядку. Для політомічних завдань складність замінюється середнім значенням завдання, а дискримінація не використовується. Якщо дані оцінені дихотомічно, то будуть відображені частоти розподілу за дистракторами, з низько- і високооціненими групами (аналіз дистракторів). Якщо дані політомічно оцінені, то буде визначена частота оцінки для кожного завдання. Для подальшого використання в окремих наборах даних зберігаються оцінки за завдання, загальний бал, а також складність і дискримінативність завдання [8].

ITEM_NUM	FVALUE	D_INDEX
1	,19	,38
2	,27	,88
3	,31	,88
4	,42	,88
5	,46	,61
6	,50	,32
7	,58	,57
8	,62	,34
9	,65	-,08
10	,73	,71
11	,73	,45
12	,73	,71
13	,73	,45
14	,77	,43
15	,77	,86
16	,81	,45
17	,81	,43
18	,81	,28
19	,85	,57
20	,92	,29
21	,92	,14
22	,96	,14
23	,96	,14
24	,96	-,13
25	,96	,14
26	,96	,14
27	1,00	,00
28	1,00	,00
29	1,00	,00
30	1,00	,00
31	1,00	,00

Рис. 2. Показники складності й дискримінативності, отримані на основі даних проведеної математичної частини тесту ТЗНК (вбірка з 26 учасників)

Обчисливши ці дані за допомогою схарактеризованого вище синтаксису SPSS, отримуємо файл з вихідними даними під назвою PD.sav. Щоб здійснити опис числових даних у текстовій формі (надати певним діапазнам значень їх рівень), скористаємось синтаксисом для пришвидшення роботи над висновками й психометричними характеристиками результатів (рис. 3).

За даними у вигляді рівнів складності будуємо діаграму частотного розподілу у відсотковому співвідношенні (рис. 4). На діаграмі завдання розподілені за 4 рівнями складності: як переконаємося, більшість завдань (77,42%) виявилася дуже легкою для цієї вибірки учасників тестування.

```

1
2
3
4
5
6 DO IF (Pvalue >= .60).
7 COMPUTE Pvalue_string = "Дуже легке".
8 ELSE IF (Pvalue > .41 and Pvalue < .60).
9 COMPUTE Pvalue_string = "Легке".
10 ELSE IF (Pvalue > .31 and Pvalue < .40).
11 COMPUTE Pvalue_string = "Оптимальне".
12 ELSE IF (Pvalue > .20 and Pvalue < .30).
13 COMPUTE Pvalue_string = "Складне".
14 ELSE.
15 COMPUTE Pvalue_string = "Дуже складне".
16 END IF.
17 EXECUTE.
    
```

Рис. 3. Синтаксис для перетворення діапазону значень складності на аналогічні рівні

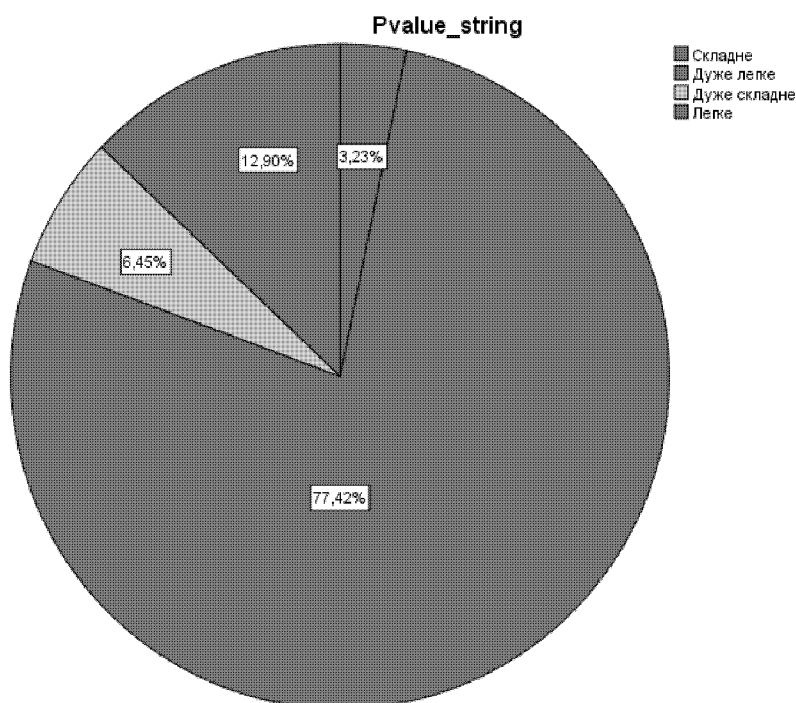


Рис. 4. Діаграма частотного розподілу завдань за рівнями складності

Коефіцієнт Альфа Кронбаха є найбільш поширеним засобом внутрішньої узгодженості. SPSS має вбудовану перевірку Альфа Кронбаха. Разом з цим, SPSS обчислює середню величину

шкали, дисперсію й кореляцію. У таблиці 2 наведено дані так, що стає зрозумілим, як зміняться статистичні показники, якщо вилючити те чи інше завдання з тесту.

Таблиця 2

Статистика співвідношення завдань з сумарним балом, отримана на основі даних проведеної математичної частини тесту ТЗНК (вибірка з 26 учасників)

	Середнє шкали за умов видалення пункту	Дисперсія шкали за умов видалення пункту	Кореляція пункту з сумарним балом	Квадрат коефіцієнта множинної кореляції	Альфа Кронбаха за умов видалення пункту
МС1	18,6786	20,522	,487	.	,840
МС2	18,6071	19,581	,678	.	,832
МС3	18,5714	19,587	,657	.	,832
МС4	18,4643	19,221	,715	.	,830
МС5	18,4286	20,402	,435	.	,842
МС6	18,3929	21,210	,255	.	,849
МС7	18,3214	20,522	,420	.	,842
МС8	18,2857	21,026	,312	.	,847
МС9	18,2500	22,417	,004	.	,858
МС10	18,1786	20,004	,624	.	,835
МС11	18,1786	20,967	,371	.	,844
МС12	18,1786	20,004	,624	.	,835
МС13	18,1786	21,115	,334	.	,845
МС14	18,1429	21,164	,344	.	,845
МС15	18,1429	19,831	,714	.	,832
МС16	18,1071	21,136	,383	.	,844
МС17	18,1071	21,433	,298	.	,846
МС18	18,1071	21,877	,173	.	,850
МС19	18,0714	20,810	,531	.	,839
МС20	18,0000	21,481	,457	.	,843
МС21	18,0000	22,148	,180	.	,849
МС22	17,9643	22,258	,206	.	,848
МС23	17,9643	22,258	,206	.	,848
МС24	17,9643	22,925	-,165	.	,854
МС25	17,9643	21,962	,375	.	,845
МС26	17,9643	22,184	,248	.	,847

Висновки. Отже, SPSS має величезні вбудовані статистичні можливості і придатний для використання через підтримку синтаксису, а тому його можна використовувати для опрацювання результатів тестування.

Результати застосування SPSS доводять, що його переваги полягають у можливості автоматизувати й налагодити роботу з опрацювання даних і швидко одержувати точну, якісну й правдиву психометричну характеристику тесту.

Список використаних джерел

1. Бахрушин В. Є. Методи аналізу даних: навч. посібник / В. Є. Бахрушин. – Запоріжжя: Класич. приват. ун-т., 2011. – 267 с.
2. Лагунова А. А. Применение программной среды SPSS для исследования данных психологических методик [Електронний ресурс] / Лагунова А. А., Муллинов Д. О., Баженов Р. И. // Психология, социология и педагогика. – 2015. – № 6 – Режим доступу: <http://psychology.snauka.ru/2015/06/5446> (24. 06. 2015).
3. Лупан І. В. Комп'ютерні статистичні пакети: навч.-метод. посібник / І. В. Лупан, О. В. Авраменко. – Кіровоград: [б. в.], 2010. – 218 с.
4. Мазорчук М. С. Оценка качества тестов на основе

References

1. Bakhrushin, V. E. (2011). *Data analysis methods: teaching guidance*. Zaporizhia: Classical Private University. [in Ukrainian]
2. Lagunova, A. A., Mullinov, D. O., Bazhenov, R. I. (2015, June 24). The use of SPSS software environment for the study of the psychological methods of data. *Psikhologiya, sotsiologiya i pedagogika*, 6. Retrieved from: <http://psychology.snauka.ru/2015/06/5446> [in Russian]
3. Lupan, I. V., Avramenko, O.V. (2010). *Computer statistical packages: teaching methodological guidance*. Kirovohrad. [in Ukrainian]
4. Mazorchuk, M. S., Bondarenko, E. O., Dobriak, V. S.

- анализа дистракторов по методу пороговых групп / М. С. Мазорчук, Е. О. Бондаренко, В. С. Добряк // *Радиоелектронні і комп'ютерні системи.* – 2013. – № 3. – С. 39–44.
5. Молчанова О. Г. Анализ систем тестирования в компьютерных обучающих комплексах / Молчанова О. Г., Морозова О. И., Чумаченко И. В. // *Система обробки інформації: зб. наук. праць.* – Харків : ХУ ПС, 2012. – Вип. 7 (105). – С. 292–296.
 6. Раков С. А. Алгоритм корректировки тестовых баллов на основе анализа сложности заданий / Раков С. А., Мазорчук М. С., Бондаренко Е. О. // *Информационные технологии в образовании.* – 2013. – № 16. – С. 49–56.
 7. Соколов А. Ю. Методы оценивания результатов тестирования в автоматизированных системах обучения [Электронный ресурс] / А. Ю. Соколов, О. Г. Молчанова // *Радиоелектронні і комп'ютерні системи.* – 2011. – № 1. – С. 117–123. – Режим доступу : http://nbuv.gov.ua/UJRN/recs_2011_1_22 (25. 03. 2016).
 8. Lei P. W. & Wu Q. CTTITEM: SAS macro and SPSS syntax for classical item analysis. *Behavior Research Methods.* – 2007. – 29(3). – PP. 527–530.
 - (2013). Tests quality assessment based on the analysis of the distractors by the method of threshold groups. *Radioelektronni i komputerni systemy*, 3, 39-44. [in Russian]
 5. Molchanova, O. G., Morozova, O. I., Chumachenko, I. V. (2012). Analysis of testing systems in computer tutoring complex. *Information processing systems: coll. of scientific works.* Kharkiv : KhU PS, Vol. 7 (105), 292-296. [in Russian]
 6. Rakov, S.A., Mazorchuk, M.S., Bondarenko, E.O. (2013). The algorithm of test scores adjustments based on the tasks complexity analysis. *Informatsionnye tekhnologii v obrazovanii*, 16, 49-56. [in Russian]
 7. Sokolov, A. U., Molchanova, O. G. (2011). Methods of evaluation of test results in the automated learning systems. *Radioelektronni i komputerni systemy*, 1, 117-123, Retrieved from : http://nbuv.gov.ua/UJRN/recs_2011_1_22. [in Russian]
 8. Lei, P. W., Wu, Q (2007). CTTITEM: SAS macro and SPSS syntax for classical item analysis. *Behavior Research Methods*, 29(3), 527-530 [in English]

Рецензент: Троїцька Т.С. – д.філософ.н., професор

Відомості про авторів:

Малежик Михайло Павлович

malez@ukr.net

Національний педагогічний університет

імені М. П. Драгоманова

вул. Пирогова, 9, м. Київ, Київська обл., 01601, Україна;

Привольнев Максим Олександрович

pryvolnev.m@gmail.com

Національний педагогічний університет

імені М. П. Драгоманова

вул. Пирогова, 9, м. Київ, Київська обл., 01601, Україна

doi: <http://dx.doi.org/10.7905/nvmdfru.v0i16.1404>

Матеріал надійшов до редакції 28. 03. 2016 р.

Прийнято до друку 29.04.2016 р.