Regression with interaction effects: determinants for choosing STEM-fields by schoolgirls

In the modern world professional orientation of a person becomes very important. On the one hand, the right professional choice can be the first step for building a successful future. On the other hand, this choice is predetermined by external factors such as the development of economical branches as well as norms, values, and rituals accompanying entering the profession. At the same time, modern men and women still do not have equal opportunities for making this choice as evidenced by a phenomenon of polarization of scientific careers on a gender basis. This inequality seems especially acute in the priority and highly paid STEM-fields, which includes natural sciences, technology, engineering, and mathematics. For example, in 2016, according to official statistics, there was only 24% of women employed in STEM-fields in Russia and Europe. And as it is known from foreign studies, the origins of this inequality should be found in school years[1].

A survey of 700 schoolchildren of the 5th, 8th, 11th grades of comprehensive schools in Moscow and Gubkin-town (Belgorod region) confirmed the presence of gender asymmetry in the preliminary choice of STEM-fields. Thus, the data obtained indicated only 35% of schoolgirls chose STEM-fields, while other gave their preference to the humanitarian and creative professions.

To search for determinants of the schoolchildren`s choice of STEM-fields, on the collected data there was built a binary logistic regression. However, the model, which included only the main effects (the predictors in their original form, without considering combinations of their categories) had a low predictive power as the value of the analog of the determination coefficient (pseudo-R2 Nagelkerke) was only 0.14. Within the framework of Research and Study Group (RSG), it was decided to build a more accurate regression model, which counts for the interaction effects. Thus, new model helped to increase the predictive power to 0.51 without adding the additional predictors.

The algorithm of the construction of the regression model with iteration effects

The non-linear form of the relation between the Y (dependent variable) and X-s (predictors) can be one of the hypothetical causes for the low explanatory power of regression models. (It should be considered that logistic regressions assume a linear relationship between the dependent (the logit) and independent variables as well). There are two ways for achieving the nonlinearity of the connection. The first one is to choose the correct nonlinear function while the other one - is to refer to the interaction effects. In this case, further work was based on the second path.

[1] *Catsambis S.* The Path to Math: Gender and Racial-Ethnic Differences in Mathematics Participation from Middle School to High School //Sociology of Education. 1994. Vol. 67. No. 3. P. 199-215. *Correll S.* Gender and the Career Choice Process: The Role of Biased Self-Assessments // American Journal of Sociology. 2001. Vol. 106. No. 6. P. 1691-1730. *Eccles J., Vida M.* Predicting Gender and Individual Differences in College Major, Career Aspirations, and Career Choice // Society for Research on Child Development.2003.

The first step of finding significant interaction effects was resorting to classification trees. One of the most suitable methods for this is ChAID. It allows finding the strongest interactions of predictors that can explain the dependent variable, which is the preliminary professional choice (the fact of choosing or not choosing by girl-students of trajectories related to the STEM). Since ChAID is based on the chi-square test, independent variables can be measured by any type of scale. Classification trees indicated 13 significant links between the predictors. It's important to work with them further.

As classification trees are intended to find iterations itself and don't include calculating model parameters and "strict" criteria of predictions quality, the next step was multidimensional dispersion analysis. This method allows to orient in the highest possible predictive ability of the chosen set of predictors regardless the form of connection between them and the dependent variable. Using dispersion analysis, from 13 iteration effects and 16 main effects the significant ones with the highest predictive ability were chosen. This procedure was held by stepwise deleting of predictors with the lowest mean square - the index that can be considered as a value of the effect's force normalized to the number of categories.

The important advantage of working with dispersion analysis is that this method, unlike regression, doesn't require the presence of variables responsible for any iterations in the base (they are created virtually). The method allows to choose the limited set of necessary iteration effects, and only they then require preparation in form of creation of additional variables. After their appropriate preparation and before the binary logistic regression start all got combinations were checked for fullness. Those, that were represented by less than 15 respondents were excluded from further analysis. This is explained by the fact that including such "rare" combinations in the model would inevitably have led it to unsustainability (irreproducibility at other data).

The entire described algorithm was dedicated to choosing iteration effects with the highest explanatory ability. They help to build the high-precision binary logistic regression and to deepen the substantive analysis. These steps allowed to increase the predictive ability more than three times (from 0.14 to 0.51) without using additional predictors. In other words, it was established that the set of variables, which didn't show the high predictive ability, in fact, had one, but only if iterations of those variables are considered.

Interpretation of interesting effects

Girls do not choose STEM if they study in the socio-economic class, their parents do not help them with homework assignments in mathematics, educational process in their school can be classified as gender-oriented[2]. On the contrary, girls tend to choose STEM-fields if they highly estimate their abilities in mathematics; they think that mathematics is a "female" discipline; they study in a class without a profile and receive help from parents in doing homework on mathematics, while the gender-directed organization of school life in school is poorly expressed.

It is important to note that the resulting interaction effects are nonlinear. This means, for example, that the gender-based organization of the school life does not itself sufficiently influence on the dependent variable. However, in combination with two other categories of variables - the class profile and getting help from the parents - the impact of the gender-focused learning organization on the choice of STEM fields is sufficient.

[2] The variable was obtained using the principal component method in the previous stages of the analysis. In this factor, the main emphasis is on the content of school lessons. For example, this included labor lessons, where boys and girls are offered to do tasks that directly reflect their gender differences.

Application

The literature used in the study of gender asymmetry in the preliminary choice of STEM fields by schoolchildren:

1. *Ambrose R., Levi L., Fennema E.* The complexity of teaching for gender equity // Multicultural and gender equity in the mathematics classroom: The gift of diversity. 1997. P. 236-242.
2. *Archer L, DeWitt J, Osborne J, Dillon J, Willis B, Wong B.* Science aspirations, capital, and family habitus: how families shape children’s engagement and identification with science *// American Educational Research Journal*. 2012. Vol. 49. No. 5. P. 881–908.
3. *Aschbacher P, Li E, Roth E.* Is science me? High school students’ identities, participation and aspirations in science, engineering, and medicine // Journal of Research in Science Teaching. 2010. Vol.47. No. 5. P. 564–82.
4. *Bandura A., et al.* Self-efficacy beliefs as shapers of children’s aspirations and career trajectories // Child Development. 2001. Vol 72. P. 187–206.
5. *Buffington C., Cerf B., Jones C., Weinberg B.* STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data Linked to the 2010 Census // American Economic Review. 2016. Vol. 106, No. 5. P. 333-338.
6. *Carlone H., Johnson A.* Understanding the science experiences of successful women of color: science identity as an analytic lens // Journal of research in science teaching. 2007. Vol. 44. No. 8. P.1187–1218.
7. *Catsambis S*. The Path to Math: Gender and Racial-Ethnic Differences in Mathematics Participation from Middle School to High School //Sociology of Education. 1994. Vol. 67. No. 3. P. 199-215.
8. *Ceci S., Williams W.* Why Aren’t More Women in Science? // Washington DC: American Psychological Association. P. 199–210.
9. *Coleman J*. Social capital in the creation of human capital // American Journal of Sociology. 1988. Vol. 94. P. 95–120.
10. *Correll S*. Gender and the Career Choice Process: The Role of Biased Self-Assessments // American Journal of Sociology. 2001. Vol. 106. No. 6. P. 1691-1730.
11. *Dee T.* How a Teacher's Gender Affects Boys and Girls. // Education Next, Fall. 2006. P. 69- 75.
12. *Eccles J.* Expectancies, values, and academic behaviors // San Francisco, CA: Achievement and achievement motives: Psychological and sociological approaches. 1983. P. 75-146.
13. *Eccles J., Vida M.* Predicting Gender and Individual Differences in College Major, Career Aspirations, and Career Choice // Society for Research on Child Development. 2003.
14. *Fennema E., Carpenter T., Jacobs V., Franke M., Levi L*. A longitudinal study of gender differences in young children's mathematical thinking // Educational Researcher. 1998. Vol.27. No. 5. P. 6-11.
15. *Fennema E., Peterson P*. Teacher - student interactions and sex-related differences in learning mathematics // Teaching and Teacher Education. 1986. Vol.2. No. 1. P. 19-42.
16. *Forgasz H., Leder G., Kloosterman P*. New Perspectives on the Gender Stereotyping of Mathematics // Mathematical Thinking and Learning. 2004. Vol.6. No. 5. P. 389-420.
17. *Gabay-Egozi L., Shavit Y., Yaish M*. Gender Differences in Fields of Study: The Role of Significant Others and Rational Choice Motivations // European Sociological Review. 2015. Vol. 31. No. 3. P. 284–297.
18. *Gonzalez H., Kuenzi J.* Science, technology, engineering, and mathematics (STEM) education: a primer. //Congressional Research Service // URL: http://www.fas.org/sgp/crs/misc/R42642.pdf. Проверено 26.11.2016.
19. *Gottfredson L., Lapan R.* Assessing gender based circumscription of occupational aspirations // Journal of Career Assessment. 1997. Vol. 5. P. 419–441.

20. *Graham M., Byars-Winston A., Hunter A., Handelsman J.* Increasing persistence of college students in STEM // Science. 2013.Vol. 341 P. 1455–1456.

21. *Herrnstein R., Murray C.* The Bell Curve: Intelligence and Class Structure in American Life // New York: Free Press. 1994.

22. *Hyde J., Fennema E., Lamon S.* Gender differences in mathematics performance // Psychological Bulletin. 1990. P. 139-155.

23. *Legewie J., DiPrete T.* High School Environments, STEM Orientations, and The Gender Gap in Science And Engineering Degrees // New-York: Columbia University. 2012.

24. *Legewie J., T. DiPrete*. The High School Environment and the Gender Gap in Science and Engineering // Sociology of Education. 2014. Vol. 87. No. 4. P. 259-280.

25. *Lynn R.* Racial and ethnic differences in intelligence in the United States on the Differential Ability Scale // Personal. Individ. Differ. 1996. Vol. 20. No. 2. P.217–273.

26. *Maccoby E., Jacklin C.* The psychology of sex differences // Stanford: Stanford University Press. 1974.

27. *McDaniel A.* The Role of Cultural Contexts in Explaining Cross-National Gender Gaps in STEM Expectations // European Sociological Review. 2016. Vol. 32. No. 1. P 122–133.

28. *Morgan S.* On the Edge of Commitment: Educational Attainment and Race in the United States // Stanford: Stanford University Press. 2005.

29. *Noddings N.* Perspectives from feminist philosophy // Educational Researcher. 1998. Vol.27. No. 5. P. 17-18.

30. OECD PISA Database // URL: http://www.oecd.org/pisa/pisaproducts/48852548.pdf. Проверено 25.11.2016.

31. OECD. The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence, PISA, OECD Publishing (2015) // URL: http://dx.doi.org/10.1787/9789264229945-en. Проверено 25.11.2016.

32. *Seligman L., Weinstock L.* The career development of 10 year olds // Elementary School Guidance and Counseling. 1991. Vol. 25. P. 172–182.

33. *Sewell W., Haller A., Portes A.* The educational and early occupational attainment process // American Sociological Review. 1969. Vol. 34. P. 82–92.

34. *Seymour E, Hewitt N.* Talking about Leaving: Why Undergraduates Leave the Sciences // Boulder, CO: Westview. 1997.

35. *Shelley J.* Gender and the Career Choice Process: The Role of Biased Self-Assessments // American Journal of Sociology. 2001. Vol. 106, No. 6. P.1691-1730.

36. *Smetackova I.* Gender Stereotypes, Performance and Identification with Math // Procedia - Social and Behavioral Sciences. 2015. Vol. 190. P. 211 – 219.

37. *Stanley J., Benbow* C. Sex differences in mathematical ability: Fact or artifact // Science. 1980. Vol. 210. No. 4475. P. 1262-1264.

38. *Tinto V.* Leaving College: Rethinking the Causes and Cures of Student Attrition // Chicago: Univ. Chicago Press. 1987.

39. *Xie Y., Fang M., Shauman K.* STEM Education // Annual Review of Sociology. 2015. No. 41. P. 331–357.

40. *Бем С.* Линзы гендера: Трансформация взглядов на проблему неравенства полов / Пер. с англ. М.: Российская политическая энциклопедия (РОССПЭН), 2004. С. 189.

41. *Дрозд В.* Гендерные различия в усвоении математики: реальность или иллюзия? // Didactics ofmathematics: Problems and Investigations. 2008. Issue # 29. P.124-134.

42. *Здравомыслова Е., Темкина А* Социальное конструирование гендера как феминистская теория // Социологический журнал. 1998. № 3-4. С.171-182.

43. *Здравомыслова Е., Темкина А.* Социальная конструкция гендера и гендерная система в России // Материалы Первой российской летней школы по женским и гендерным исследованиям «Валдай 96». М., 1997. С.84-85.

44. *Киммел М.* Гендерное общество / Пер. с англ. М.: Российская политическая энциклопедия (РОССПЭН), 2006. С. 464.

45. *Малышева М.* Естественные и технические науки для женщин в XXI веке // Народонаселение. 2016. № 3. С. 76-85.

46. *Савинская О.* Гендерное равенство в STEM-программах дошкольного образования как фактор успешного технологического развития России // Женщина в российском обществе. 2016. №3. С. 16-24.

47. Федеральная служба государственной статистики. Сборник «Женщины и мужчины России». 2014 г. //URL: http://www.gks.ru/bgd/regl/b14_50/Main.htm. Проверено 20.11.2016.

48. *Хасбулатова О.* Гендерные аспекты развития STEM-образования в России // Женщина в российском обществе. 2016. №3. С. 3-15.

49. *Шумилова О.* Политическое поведение женщин в условиях реформирования современного российского общества: Монография / Под общей ред. д-ра полит. наук, проф. И.А. Батаниной. Тула: ТулГУ, 2005. С. 182.