

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**

Carla Fiorioli

**QUANTITATIVE ANALYSIS OF STUDENTS'
PERFORMANCE IN THE BRAZILIAN HIGH SCHOOL
EXAM**

Porto Alegre

2019

Carla Fiorioli

**QUANTITATIVE ANALYSIS OF STUDENTS' PERFORMANCE IN THE
BRAZILIAN HIGH SCHOOL EXAM**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Mestre em Engenharia de Produção, modalidade Acadêmica.

Orientador: Flávio Sanson Fogliatto, *Ph.D.*

Porto Alegre

2019

Carla Fiorioli

**QUANTITATIVE ANALYSIS OF STUDENTS' PERFORMANCE IN THE
BRAZILIAN HIGH SCHOOL EXAM**

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia de Produção na modalidade Acadêmica e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul.

Flávio Sanson Fogliatto, *Ph.D.*

Orientador PPGEP/UFRGS

Flávio Sanson Fogliatto, *Ph.D.*

Coordenador PPGEP/UFRGS

Banca Examinadora:

Professor José Luís Duarte Ribeiro, Dr. (PPGEP/UFRGS)

Professor Michel José Anzanello, *Ph.D.* (PPGEP/UFRGS)

Priscila Goergen Brust Renck, *Ph.D.* (PPGEP/UFRGS)

Dedico essa dissertação à minha família e em especial aos meus pais José Carlos e Rosana, à minha irmã Thais e à minha sobrinha e afilhada Kiara.

AGRADECIMENTOS

Em primeiro lugar, agradeço aos meus pais, José Carlos e Rosana, pelos ensinamentos, pelo carinho, pelo apoio incondicional, por sempre acreditarem na minha capacidade e pelo papel decisivo que tiveram na conclusão desta dissertação.

Agradeço também ao meu professor e orientador, Flávio Sanson Fogliatto, por sua dedicação e por sua valiosa orientação, também imprescindíveis à conclusão deste trabalho.

À minha irmã, Thais, por sempre acreditar em mim, pelas palavras de apoio nos momentos mais difíceis e pelo carinho. À minha sobrinha e afilhada Kiara, pelos sorrisos, pela leveza e pelos momentos de descontração proporcionados. À minha tia e madrinha Neiva e à minha tia Graça pelo carinho, pelo apoio e pela compreensão. À toda minha família pelo carinho e pela compreensão mesmo na minha ausência.

A todos os meus amigos, em especial às minhas amigas Camila, Joana e Marina, e ao meu amigo Pedro pela compreensão, pelos momentos de descontração e pelas palavras de incentivo. Aos amigos Mateus Hexsel e Gabriela Cardoso pelo apoio e pelas intervenções em momentos fundamentais.

Aos meus colegas de curso pelo apoio e pelo companheirismo, em especial à amiga Raísa e ao amigo Bernardo pelas conversas e pelas palavras de incentivo e à colega Gabrielli pelo ágil apoio em um momento fundamental para a conclusão deste trabalho.

Aos professores, coordenação e equipe do PPGEP, por todo suporte e dedicação. Ao CNPq pelo auxílio financeiro.

Agradeço também a todos que passaram pela minha vida e que me proporcionaram aprendizados e me fizeram crescer.

Por fim, agradeço pela energia universal e sagrada que me proporcionou a força necessária para trilhar esse caminho.

*“Querer ser livre é também querer livres os
outros.”*

Simone de Beauvoir

ABSTRACT

This dissertation presents an analysis of the relationship between sociodemographic attributes and students' performance in the Brazilian High School Exam (ENEM). The main objective is to analyze the impact of sociodemographic variables on students' performance by modeling their score in the five knowledge areas evaluated in ENEM as a function of variables that characterize these students. The method adopted to perform the analysis was an adaptation of the CRISP-DM (Cross-Industry Standard Process for Data Mining) method, which is implemented in five phases: (i) data/environment understanding, (ii) data preparation, (iii) modeling, (iv) evaluation/analysis, and (v) discussion. In all models obtained, the variable that most explained the variance in students' performance was a dummy variable associated with the type of school attended by the student: those who attended only private schools without a scholarship had an advantage in test scores. A dummy variable related to race was also retained in all models: auto declared white students had an advantage in scores. The sex-related effect varied depending on the area of knowledge analyzed. The most positive effect for males occurred in the mathematics knowledge area, while the most negative effect occurred in the essay. Other variables such as students' fathers' and mothers' level of education, fathers' occupation, and the ownership of a computer were included in all models. Models obtained yielded an average variance explained of 17.90%, which is consistent to what is observed in other studies of the same nature, and suggests that the method employed is suitable for this type of analysis.

Keywords: Assessment exams, ENEM, sociodemographic attributes, data mining, partial least squares regression.

RESUMO

Esta dissertação apresenta uma análise quantitativa do desempenho dos estudantes no Exame Nacional do Ensino Médio (ENEM) em relação a variáveis sociodemográficas. O objetivo deste trabalho é a análise do impacto das variáveis sociodemográficas no desempenho dos estudantes através da modelagem deste desempenho nas 5 áreas de conhecimento avaliadas no ENEM como função das variáveis que caracterizam esses estudantes. Para a realização da análise, utilizou-se uma versão adaptada do método CRISP-DM (*Cross-Industry Standard Process for Data Mining*) implementado em cinco fases: (i) compreensão dos dados/ambiente, (ii) preparação dos dados, (iii) modelagem, (iv) avaliação/análise e (v) discussão. Em todos os modelos gerados, a variável que melhor explica a variação no desempenho dos estudantes é a variável *dummy* associada ao tipo de escola frequentada pelo aluno; aqueles que frequentavam apenas escolas privadas sem bolsa de estudos tiveram vantagem nos escores. A variável *dummy* relacionada à raça também foi mantida em todos os modelos: os estudantes auto declarados brancos tiveram vantagem nos escores. O efeito relacionado ao gênero variou dependendo da área de conhecimento analisada. O efeito mais positivo para o sexo masculino ocorreu na área de conhecimento de matemática, enquanto o efeito mais negativo ocorreu na redação. Outras variáveis, como o nível de escolaridade dos pais e mães dos estudantes, a ocupação dos pais e a posse de um computador foram incluídas em todos os modelos. A modelagem desenvolvida nesta dissertação explica, em média, 17,90% da variância do desempenho dos estudantes no ENEM, o que é consistente com os resultados obtidos em estudos de mesma natureza. Esta condição sugere que o método utilizado é adequado à realização deste tipo de análise.

Palavras-chave: Sistemas de avaliação, ENEM, atributos sociodemográficos, mineração de dados, regressão por mínimos quadrados parciais.

CONTENTS

ABSTRACT	7
RESUMO	8
1. INTRODUÇÃO	10
1.1. TEMA E OBJETIVOS	11
1.2. JUSTIFICATIVA	11
1.3. DELIMITAÇÕES DO TRABALHO	12
1.4. ESTRUTURA DO TRABALHO	13
2. BACKGROUND	14
2.1. ASSESSMENT EXAMS.....	14
2.1.1. PISA	16
2.1.2. ENEM	17
2.2. PARTIAL LEAST SQUARES (PLS).....	18
3. DATA AND METHODS	22
3.1. DATA DESCRIPTION/EXPLORATION.....	23
3.2. DATA PREPARATION.....	23
3.3. MODELING	24
3.4. EVALUATION	25
4. RESULTS AND DISCUSSIONS	28
4.1. DESCRIPTIVE ANALYSIS.....	28
4.2. MODELING.....	34
4.3. FINAL MODELS AND COEFFICIENT ANALYSES	36
4.3.1. ESSAY	39
4.3.2. LANGUAGES AND CODES.....	44
4.3.3. HUMAN SCIENCES.....	47
4.3.4. NATURAL SCIENCES.....	50
4.3.5. MATHEMATICS	52
4.3.6. COMPARISON OF MODELS.....	55
4.3.7. DISCARDED INDEPENDENT VARIABLES.....	58
5. FINAL CONSIDERATIONS	60
5.1. CONCLUSION.....	60
5.2. SUGGESTIONS FOR FURTHER STUDIES	62
REFERENCES	63
APPENDIX A – VARIABLE DESCRIPTION	69
APPENDIX B – SHAPIRO-WILK AND D’AGOSTINO-PEARSON’S NORMALITY TESTS RESULTS	75
APPENDIX C – MANN-WHITNEY TEST RESULTS	79
APPENDIX D – SMC AND VIP VALUES	88

1. INTRODUÇÃO

Sistemas de avaliação têm como um dos propósitos a promoção da igualdade na educação (Airasian, 1988). Desde o início do século 20, o *British School Certificate* passou a ser o documento oficial que atesta a conclusão do ensino médio e permite que os estudantes concorram por vagas em universidades Britânicas (Broadfoot, 2012). Ao longo do tempo, países como os Estados Unidos da América, a Austrália, a Nova Zelândia, e o próprio Reino Unido passaram a associar o seu desenvolvimento econômico com o desempenho educacional dos seus estudantes, e certificados como esses passaram a ser utilizados para monitorar e melhorar seus sistemas educacionais (Brown e Lauder, 1996).

Existem, no entanto, pontos positivos e negativos na utilização de sistemas de avaliação, tanto do ponto de vista de certificação e seleção, como do ponto de vista de utilização para desenvolvimento de políticas sociais. A natureza objetiva das avaliações fornece não apenas um meio de efetivamente comparar indivíduos, mas também induz um senso de justiça e representa a ordem e o controle (Airasian, 1988). Todavia, existem argumentos a favor da subjetividade do conhecimento que corroboram a teoria de que se somos seres sociais que desenvolvem percepções de valores e experiências; assim, as avaliações projetadas por seres sociais carregam o viés de dois grupos: aqueles que as desenvolveram e aqueles que as avaliam. Os vieses transferidos para esses sistemas tendem a sistematizar e reforçar a estratificação social (Gipps, 1999).

Independentemente das desvantagens dos sistemas de avaliação educacionais, há uma consciência de que o desenvolvimento econômico das nações depende do desenvolvimento educacional de sua população (Eckstein, 1996). A Alemanha, a França e o Japão, considerando prós e contras, decidiram sistematizar os procedimentos de avaliação e centralizar os programas nacionais de educação; esses países estão geralmente posicionados no topo dos rankings de educação (Green, 1997).

No Brasil, o ENEM (Exame Nacional do Ensino Médio) é o Exame Nacional do Ensino Médio anual realizado por estudantes. Foi criado em 1998 e é administrado pelo Instituto Nacional de Estudos e Pesquisas Educacionais (INEP) com o objetivo principal de avaliar o desempenho dos alunos na conclusão do Ensino Médio. Desde 2009, o exame tem sido amplamente adotado como uma nota principal ou complementar para a entrada em universidades públicas. A participação no exame aumentou de cerca de 160.000 estudantes em 1998 para mais de 4 milhões em 2008. Após a adoção pelas universidades, a participação

aumentou ainda mais, alcançando mais de 8,5 milhões de estudantes em 2016. O ENEM não é e nunca foi um exame obrigatório (INEP, 2017).

Independentemente dos pontos de vista divergentes em relação aos exames de avaliação, o ENEM é realizado por milhões de estudantes, sendo uma fonte prolífica de informações para entender o que está impactando o desempenho dos estudantes brasileiros (Vahdat et al., 2015). Seguindo experiências internacionais bem-sucedidas, políticas públicas e currículos centralizados/padronizados poderiam ser desenvolvidos com base no conhecimento extraído desse tipo de dado (Runci et al., 2017), permitindo o desenvolvimento de melhores estratégias educacionais (Vahdat et al., 2015; Connelly et al., 2016; De Rosa, 2017).

1.1. TEMA E OBJETIVOS

De acordo com a contextualização anteriormente exposta, propõe-se o seguinte tema: a análise quantitativa do desempenho dos alunos no Exame Nacional do Ensino Médio (ENEM) em relação a variáveis sociodemográficas.

O objetivo geral deste trabalho é a análise do impacto das variáveis sociodemográficas no desempenho dos estudantes através da modelagem deste desempenho nas 5 áreas de conhecimento avaliadas no ENEM, como função das variáveis que caracterizam esses estudantes.

Os objetivos específicos são: (i) a análise dos coeficientes gerados pelos modelos, (ii) as discussões acerca da análise dos coeficientes, e (iii) a análise comparativa dos resultados obtidos com resultados da literatura.

1.2. JUSTIFICATIVA

Como mencionado anteriormente, o ENEM tem sido amplamente adotado pelas universidades públicas como um score complementar e/ou principal e os microdados fornecidos pelo INEP (2017) têm uma riqueza de detalhes e podem ser explorados com eficiência usando ferramentas de mineração de dados. Eckstein (1996) argumenta que as nações estão conscientes de que seu desenvolvimento econômico depende do desenvolvimento educacional de sua população. Brown e Lauder (1996) acrescentam ainda que diversos países têm utilizado os resultados de exames como o ENEM para monitorar e melhorar seus sistemas educacionais.

Experiências internacionais mostram que o estudo dos microdados pode trazer subsídios para o desenvolvimento do plano educacional e de políticas públicas relacionadas à educação (Runci et al., 2017), permitindo a elaboração de melhores estratégias educacionais (Vahdat et al., 2015; Connelly et al., 2016; De Rosa, 2017). Desta forma, é de interesse do INEP (2017) e socialmente relevante que os dados do ENEM sejam objeto de estudos que contemplem diversas abordagens.

A julgar pela pesquisa bibliográfica conduzida neste trabalho, este é o primeiro estudo a usar técnicas de regressão de segunda geração para mapear e analisar as relações entre os atributos sociodemográficos e o desempenho dos estudantes usando o conjunto de dados do ENEM.

1.3. DELIMITAÇÕES DO TRABALHO

Este trabalho foi realizado tendo como referência a base de dados do ENEM do ano de 2017. O estudo contempla os três estados da região Sul (Rio Grande do Sul, Santa Catarina e Paraná). O Brasil é um país amplo, com diferentes culturas e características. Os três estados foram selecionados devido às semelhanças dessas populações e suas diferenças com relação a outros estados e regiões. Esta foi uma abordagem necessária, uma vez que a análise de todo o conjunto de dados poderia resultar em conclusões pouco generalizáveis. Menezes-Filho (2007) e Viggiano e Mattos (2013) apresentam evidências do efeito dessas diferenças. Em função disso, há a oportunidade de aplicação do mesmo método também para as demais regiões brasileiras.

Em relação aos dados utilizados nesse estudo, não foram explorados os aspectos relacionados à localização dos alunos dentro da região e de cada estado, à qualidade individual das escolas que cada aluno frequentou, bem como características individuais de saúde, deficiências ou necessidades especiais. Essas variáveis poderiam servir de subsídio para outros estudos que avaliem outras dimensões do desempenho dos alunos, não contempladas no presente estudo.

O ENEM pode ser realizado por pessoas de qualquer idade. Nesse estudo, foi tomada a decisão de se manter todos os estudantes na análise, independentemente de idade. Estudos similares poderiam ser desenvolvidos utilizando-se somente as observações cujos alunos se enquadram em alguma faixa etária específica.

Em relação aos métodos e ferramentas utilizadas durante a preparação dos dados e da geração e avaliação dos modelos, estes foram escolhidos conforme a sua adaptabilidade para o

caso em questão, mas não necessariamente são os únicos existentes. Esse estudo não tem como pretensão cobrir todos os métodos e ferramentas possíveis para cada etapa do processo. No entanto, há a responsabilidade de utilizar métodos e ferramentas adequadas para cada situação.

1.4. ESTRUTURA DO TRABALHO

Este trabalho está estruturado em 5 capítulos. O capítulo 1 apresenta uma visão geral do trabalho, apresentando seu tema e objetivos, justificativa, delimitações e a sua estrutura. O capítulo 2 traz uma revisão da literatura acerca do tema proposto e dos métodos utilizados no desenvolvimento do trabalho. O capítulo 3 apresenta a metodologia e o detalhamento das etapas utilizadas no desenvolvimento do modelo enquanto que o capítulo 4 traz uma discussão acerca dos resultados obtidos. Por fim, o capítulo 5 faz o fechamento do trabalho, apresentando as principais conclusões obtidas, bem como sugestões para trabalhos futuros.

REFERENCES

- Airasian, P. W. (1988). Measurement driven instruction: A closer look. *Educational Measurement: Issues and Practice*, 7(4), 6-11.
- Alves, M. T. G., & Soares, J. F. (2009). Medidas de nível socioeconômico em pesquisas sociais: uma aplicação aos dados de uma pesquisa educacional. *Opinião Pública*, 15(1), 1-30.
- Arai, N. H., & Matsuzaki, T. (2014). The impact of ai on education—can a robot get into the university of Tokyo. In *Proc. ICCE* (pp. 1034-1042).
- Broadfoot, P. (1996). *Education, assessment and society: A sociological analysis*. Open University Pres.
- Broadfoot, P. (2012). *Assessment, schools and society*. Routledge.
- Brown, P., & Lauder, H. (1996). Education, globalization and economic development. *Journal of Education Policy*, 11(1), 1-25.
- Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (2013). Na vida dez; na escola zero: os contextos culturais da aprendizagem da matemática. *Cadernos de pesquisa*, (42), 79-86.
- Cassel, C., Hackl, P., & Westlund, A. H. (1999). Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of Applied Statistics*, 26(4), 435-446.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American journal of sociology*, 94, S95-S120.
- Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59, 1-12.
- D'Agostino, R. B., Belanger, A., & D'Agostino Jr, R. B. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4), 316-321.
- De Lange, M., Dronkers, J., & Wolbers, M. H. (2014). Single-parent family forms and children's educational performance in a comparative perspective: Effects of school's share of single-parent families. *School Effectiveness and School Improvement*, 25(3), 329-350.
- de Oliveira Barbosa, M. L. (2009). *Desigualdade e desempenho: uma introdução à sociologia da escola brasileira*. Argumentum.

de Oliveira, I. S. V., da Silva, M. V. B., & de Siqueira, L. B. O. (2010). Determinantes do desempenho dos estudantes no vestibular da Universidade Federal da Paraíba. *Revista Economia e Desenvolvimento*, 7(2).

De Rosa, R. (2017). *Governing by Data: Some Considerations on the Role of Learning Analytics in Education*. In *Data Science and Social Research* (pp. 67-77). Springer, Cham.

Duncan, O. D. (1961). A socioeconomic index for all occupations. *Class: Critical Concepts*, 1, 388-426.

Duru-Bellat, M. & Kieffer, A. (2008). From the baccalauréat to higher education in France: Shifting inequalities. *Population*, 63(1), 119-154.

Eckstein, M. A. (1996). A comparative assessment of assessment. 233-240.

Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (2010). *Handbook of partial least squares: Concepts, methods and applications*. Heidelberg, Dordrecht, London, New York: Springer.

Evans, J. R., & Olson, D. L. (2003). *Statistics, data analysis, and decision modeling*. Prentice Hall.

Fernandes, D. C. (2004). Race, socioeconomic development and the educational stratification process in Brazil. *Research in social stratification and mobility*, 22, 365-422.

Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, 440-452.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, No. 10). New York, NY, USA:: Springer series in statistics.

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica chimica acta*, 185, 1-17.

Gipps, C. (1999). Chapter 10: Socio-cultural aspects of assessment. *Review of research in education*, 24(1), 355-392.

Granato, D., & Ares, G. (Eds.). (2014). *Mathematical and statistical methods in food science and technology*. John Wiley & Sons. 137-149.

Green, A. (1997). Education, globalization and the nation state. In *Education, Globalization and the Nation State* (pp. 130-186). Palgrave Macmillan, London.

Guest, M. (2008). Japanese university entrance examinations: What teachers should know. *LANGUAGE TEACHER-KYOTO-JALT-*, 32(2), 15.

Gujarati, D. N. (2009). *Basic econometrics*. Tata McGraw-Hill Education.

Haenlein, M., & Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding statistics*, 3(4), 283-297.

Hair, N. L., Hanson, J. L., Wolfe, B. L., & Pollak, S. D. (2015). Association of child poverty, brain development, and academic achievement. *JAMA pediatrics*, 169(9), 822-829.

Hardy, M. A. (1993). *Regression with dummy variables* (Vol. 93). Sage.

Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., & Baird, J. A. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. *Scandinavian Journal of Educational Research*, 62(3), 333-353.

IBGE: Instituto Brasileiro de Geografia e Estatística. *Síntese de indicadores sociais: uma análise das condições de vida da população brasileira*, 2017.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2017). *Microdata from Enem 2017*. <http://portal.inep.gov.br/microdados>. Accessed 31 July 2018.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018). <http://ideb.inep.gov.br/resultado/>. Accessed 27 February 2019.

Jerrim, J. (2012). The Socio-Economic Gradient in Teenagers' Reading Skills: How Does England Compare with Other Countries? *Fiscal Studies*, 33(2), 159-184.

Keeves, J. P. (1994). *National examinations: design, procedures and reporting* (Vol. 50). Unesco.

Lafontaine, D., & Monseur, C. (2009). Gender Gap in Comparative Studies of Reading Comprehension: to what extent do the test characteristics make a difference? *European Educational Research Journal*, 8(1), 69-79.

Lietz, P. (2006). A meta-analysis of gender differences in reading achievement at the secondary school level. *Studies in Educational Evaluation*, 32(4), 317-344.

Liland, K. H., Mehmood, T., Sæbø, S. (2017). Package 'plsVarSel'. 23p.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50-60.

Martins, L., & Veiga, P. (2010). Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities? *Economics of Education Review*, 29(6), 1016-1033.

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American psychologist*, 53(2), 185.

Menezes-Filho, N. A. (2007). Os determinantes do desempenho escolar do Brasil (pp. 1-31). IFB.

Mevik, B. H., Wehrens, R. & Liland, K. H. (2016). Package 'pls'. 59p.

Nonoyama-Tarumi, Y. (2008). Cross-national estimates of the effects of family background on student achievement: A sensitivity analysis. *International Review of Education*, 54(1), 57-82.

OECD. (2000). *Measuring Student Knowledge and Skills: The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy*. Paris: OECD.

Perry, L. B., & McConney, A. (2010). Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record*, 112(4), 1137-1162.

Pilz, M. (2009). After Abitur, First an Apprenticeship and then University? Why German Abitur Holders Are Taking Vocational Training in the Financial Services Sector. *European Journal of vocational training*, 46(1), 41-65.

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. " O'Reilly Media, Inc."

Randler, C., & Frech, D. (2006). Correlation between morningness–eveningness and final school leaving exams. *Biological Rhythm Research*, 37(3), 233-239.

Reichelt, M. (1997). Writing instruction at the German Gymnasium: A 13th-grade English class writes the Abitur. *Journal of Second Language Writing*, 6(3), 265-291.

Runci, M. C., Di Bella, G., & Cuppone, F. (2017). Integrated Education Microdata to Support Statistics Production. In *Data Science and Social Research* (pp. 283-290). Springer, Cham.

Sampaio, B., & Guimarães, J. (2009). Diferenças de eficiência entre ensino público e privado no Brasil. *Economia Aplicada*, 13(1), 45-68.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591-611.

Shavit, Y. (Ed.). (2007). *Stratification in higher education: A comparative study*. Stanford University Press.

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing*, 5(4), 13-22.

Silva, E. M. D., & Araújo, D. L. D. (2009). University entrance exam: the washback effect of textual genre knowledge. *Trabalhos em Linguística Aplicada*, 48(1), 133-152.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453.

Stearns, B., Rangel, F., Rangel, F., de Faria, F. F., Oliveira, J., & Ramos, A. A. D. S. (2017). Scholar Performance Prediction using Boosted Regression Trees Techniques. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*. Citeseer.

Steinmayr, R., & Spinath, B. (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation? *European Journal of Personality: Published for the European Association of Personality Psychology*, 22(3), 185-209.

Stewart, E. B. (2008). School structural characteristics, student effort, peer associations, and parental involvement: The influence of school-and individual-level factors on academic achievement. *Education and urban society*, 40(2), 179-204.

Tran, T. N., Afanador, N. L., Buydens, L. M., & Blanchet, L. (2014). Interpretation of variable importance in partial least squares with significance multivariate correlation (SMC). *Chemometrics and Intelligent Laboratory Systems*, 138, 153-160.

Tucker-Drob, E. M. (2013). How many pathways underlie socioeconomic differences in the development of cognition and achievement? *Learning and Individual Differences*, 25, 12-20.

Vahdat, M., Ghio, A., Oneto, L., Anguita, D., Funk, M., & Rauterberg, M. (2015). Advances in learning analytics and educational data mining. *Proc. of ESANN2015*, 297-306.

Viggiano, E., & Mattos, C. (2013). O desempenho de estudantes no Enem 2010 em diferentes regiões brasileiras. *Revista Brasileira de Estudos Pedagógicos*, 94(237).

Watanabe, Y. (2013). The National Center Test for University Admissions. *Language Testing*, 30(4), 565-573.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological bulletin*, 91(3), 461.

Wold, S., Ruhe, A., Wold, H., & Dunn, III, W. J. (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, 5(3), 735-743.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2), 109-130.

Zhou, Y., Zhu, Y., & Leung, S. W. (2014). PLS-Fraily Model for Cancer Survival Analysis Based on Gene Expression Profiles. In *International Conference on Partial Least Squares and Related Methods* (pp. 189-199). Springer, Cham.

APPENDIX A – VARIABLE DESCRIPTION

INDEPENDENT VARIABLES			
VARIABLE	DESCRIPTION	CATEGORY	CATEGORY DESCRIPTION
AGE	Student's age	-	-
SEX	Student's sex	F* M	Female Male
RACE	Student's auto declared race	A* B C D E F	Black Mixed White Yellow Indigenous Not declared
LG	Student's foreign language option	S* E	Spanish English
SC_TP	Student's type of school attended	A* B C D E	Only public school Part in public school and part in private school with no scholarship Part in public school and part in private school with scholarship Only private school with no scholarship Only private school with scholarship
HS_ST	Situation of student's High School diploma	A* B C D E F G H I J K L M N	Never completed High School and it is not attending classes to complete Still have more than one year to finish High School Is in the last year of High School Completed High School by 2016 Completed High School by 2015 Completed High School by 2014 Completed High School by 2013 Completed High School by 2012 Completed High School by 2011 Completed High School by 2010 Completed High School by 2009 Completed High School by 2008 Completed High School by 2007 Completed High School before 2007

INDEPENDENT VARIABLES			
VARIABLE	DESCRIPTION	CATEGORY	CATEGORY DESCRIPTION
IN	Family's monthly income	A*	Up to R\$ 1,405.50 (1.5 minimum wages)
		B	From R\$ 1,405.51 up to R\$ 1,874.00 (from 1.5 to 2 minimum wages)
		C	From R\$ 1,874.01 up to R\$ 2,342.50 (from 2 to 2.5 minimum wages)
		D	From R\$ 2,342.51 up to R\$ 2,811.00 (from 2.5 to 3 minimum wages)
		E	From R\$ 2,811.01 up to R\$ 3,748.00 (from 3 to 4 minimum wages)
		F	From R\$ 3,748.01 up to R\$ 4,685.00 (from 4 to 5 minimum wages)
		G	From R\$ 4,685.01 up to R\$ 5,622.00 (from 5 to 6 minimum wages)
		H	From R\$ 5,622.01 up to R\$ 6,559.00 (from 6 to 7 minimum wages)
		I	From R\$ 6,559.01 up to R\$ 7,496.00 (from 7 to 8 minimum wages)
		J	From R\$ 7,496.01 up to R\$ 8,433.00 (from 8 to 9 minimum wages)
		K	From R\$ 8,433.01 up to R\$ 9,370.00 (from 9 to 10 minimum wages)
		L	From R\$ 9,370.01 up to R\$ 11,244.00 (from 10 to 12 minimum wages)
		M	From R\$ 11,244.01 up to R\$ 14,055.00 (from 12 to 15 minimum wages)
		N	From R\$ 14,055.01 up to R\$ 18,740.00 (from 15 to 20 minimum wages)
		O	More than R\$ 18,740.00 (20 minimum wages)
		ST_FA	Years of schooling of student's father
B	Studied but did not complete primary education		
C	Completed primary education but did not complete lower secondary education		
D	Completed lower secondary education but did not complete High School		
E	Completed High School but did not complete tertiary education		
F	Completed tertiary education but did not complete graduate studies		
G	Completed graduate studies		
H	Not informed		
ST_MO	Years of schooling of student's mother	A*	Never studied
		B	Studied but did not complete primary education
		C	Completed primary education but did not complete lower secondary education
		D	Completed lower secondary education but did not complete High School
		E	Completed High School but did not complete tertiary education
		F	Completed tertiary education but did not complete graduate studies
		G	Completed graduate studies
		H	Not informed

INDEPENDENT VARIABLES			
VARIABLE	DESCRIPTION	CATEGORY	
OC_FA	Student's father occupation	A*	Farmer, livestock farmer (cattle, pigs, chickens, sheep, horses, etc.), beekeeper, fisherman, lumberjack, rubber tapper, extractivist
		B	Housekeeper, elderly caregiver, nanny, cook (in private homes), private driver, gardener, janitor, guard, porter, postman, administrative assistant, receptionist, bricklayer, repositories of merchandise
		C	Baker, industrial or restaurant cook, cobbler, dressmaker, jeweler, mechanic, machine operator, welder, factory worker, mining worker, painter, electrician, plumber, driver, truck driver, taxi driver
		D	Teacher (primary or secondary education, language, music, arts etc.), technician (nursing, accounting, electronics etc.), police officer, low military officer (soldier, corporal, sergeant), supervisor, manager, master builder, pastor, micro entrepreneur (owner of a company with less than 10 employees), small trader, small landowner, self-employed or self-employed
		E	Doctor, engineer, dentist, psychologist, economist, lawyer, judge, promoter, defender, delegate, lieutenant, captain, colonel, university professor, director in public or private companies, politician, owner of companies with more than 10 employees
		F	Not informed
OC_MO	Student's mother occupation	A*	Farmer, livestock farmer (cattle, pigs, chickens, sheep, horses, etc.), beekeeper, fisherwoman, lumberjack, rubber tapper, extractivist
		B	Housekeeper, elderly caregiver, nanny, cook (in private homes), private driver, gardener, janitor, guard, porter, postman, administrative assistant, receptionist, bricklayer, repositories of merchandise
		C	Baker, industrial or restaurant cook, cobbler, dressmaker, jeweler, mechanic, machine operator, welder, factory worker, mining worker, painter, electrician, plumber, driver, truck driver, taxi driver
		D	Teacher (primary or secondary education, language, music, arts etc.), technician (nursing, accounting, electronics etc.), police officer, low military officer (soldier, corporal, sergeant), supervisor, manager, master builder, pastor, micro entrepreneur (owner of a company with less than 10 employees), small trader, small landowner, self-employed or self-employed
		E	Doctor, engineer, dentist, psychologist, economist, lawyer, judge, promoter, defender, delegate, lieutenant, captain, colonel, university professor, director in public or private companies, politician, owner of companies with more than 10 employees
		F	Not informed

INDEPENDENT VARIABLES		
VARIABLE	DESCRIPTION	CATEGORY DESCRIPTION
RES	Quantity of residents in the student's house (including the student)	A* 1
		B 2
		C 3
		D 4
		E 5
		F 6
		G 7
		H 8
		I 9
		J 10 or more
CAR	If there are one or more cars in the student's house	N* No
		Y Yes
MOTO	If there are one or more motorcycles in the student's house	N* No
		Y Yes
CELL	Quantity of cell phones in the student's residence	A* None or one
		B Two or more
PC	If there are one or more computers in the student's house	N* No
		Y Yes
INT	If there is access to the internet in the student's house	N* No
		Y Yes
CB_TV	If there is cable TV installed in the student's house	N* No
		Y Yes

INDEPENDENT VARIABLES			
VARIABLE	DESCRIPTION	CATEGORY	CATEGORY DESCRIPTION
TV	Quantity of televisions in the student's house	A* B	None or one Two or more
DVD	If there is a DVD player in the student's house	N* Y	No Yes
TP	If there is a fixed telephone installed in the student's house	N* Y	No Yes
HK	If a housekeeper works in the student's house	N* Y	No Yes
BATH	Quantity of bathrooms in the student's house	A* B	None or one Two or more
BED	Quantity of bedrooms in the student's house	A* B	None or one Two or more
MW	If there are one or more microwaves in the student's house	N* Y	No Yes
FGE	Quantity of fridges in the student's house	A* B	None or one Two or more
FZR	If there are one or more freezers in the student's house	N* Y	No Yes

INDEPENDENT VARIABLES			
VARIABLE	DESCRIPTION	CATEGORY	CATEGORY DESCRIPTION
WA_MA	If there are one or more washing machines in the student's house	N* Y	No Yes
DR_MA	If there are one or more drying machines in the student's house	N* Y	No Yes
DI_WA	If there are one or more dish washers in the student's house	N* Y	No Yes
VA_CL	If there is a vacuum cleaner in the student's house	N* Y	No Yes

DEPENDENT VARIABLES			
VARIABLE	DESCRIPTION	CATEGORY	CATEGORY DESCRIPTION
ES_SC	Student's score for the Essay	-	-
LC_SC	Student's score for Languages and Codes	-	-
HS_SC	Student's score for Human Sciences	-	-
NS_SC	Student's score for Natural Sciences	-	-
MA_SC	Student's score for Mathematics	-	-

* Variable categories that were used as the control group. No dummy variables were created from these categories.

APPENDIX B – SHAPIRO-WILK AND D’AGOSTINO-PEARSON’S NORMALITY TESTS RESULTS

	ES			LC			HS					
	SHAPIRO-WILK		D’AGOSTINO-PEARSON’S	SHAPIRO-WILK		D’AGOSTINO-PEARSON’S	SHAPIRO-WILK		D’AGOSTINO-PEARSON’S			
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value		
ST_FA												
A	0.992	0.000	80.069	0.000	0.986	0.000	302.227	0.000	0.993	0.000	189.628	0.000
B	0.989	0.000	1587.714	0.000	0.987	0.000	2850.620	0.000	0.992	0.000	1882.769	0.000
C	0.989	0.000	1522.840	0.000	0.987	0.000	2918.597	0.000	0.992	0.000	1464.435	0.000
D	0.987	0.000	1656.136	0.000	0.987	0.000	2547.816	0.000	0.991	0.000	1245.872	0.000
E	0.988	0.000	2781.047	0.000	0.983	0.000	6619.970	0.000	0.989	0.000	2479.229	0.000
F	0.992	0.000	271.168	0.000	0.973	0.000	3250.026	0.000	0.980	0.000	1536.390	0.000
G	0.993	0.000	113.528	0.000	0.963	0.000	3009.463	0.000	0.972	0.000	1673.757	0.000
H	0.990	0.000	597.712	0.000	0.987	0.000	1188.277	0.000	0.991	0.000	730.272	0.000
ST_MO												
A	0.991	0.000	80.372	0.000	0.985	0.000	283.431	0.000	0.992	0.000	163.693	0.000
B	0.990	0.000	974.942	0.000	0.985	0.000	2519.229	0.000	0.991	0.000	1496.603	0.000
C	0.989	0.000	1356.555	0.000	0.986	0.000	2871.110	0.000	0.991	0.000	1577.737	0.000
D	0.989	0.000	1532.405	0.000	0.986	0.000	2879.694	0.000	0.992	0.000	1342.734	0.000
E	0.989	0.000	3235.965	0.000	0.985	0.000	6437.257	0.000	0.990	0.000	2734.623	0.000
F	0.991	0.000	391.603	0.000	0.976	0.000	3324.687	0.000	0.982	0.000	1488.920	0.000
G	0.992	0.000	269.375	0.000	0.973	0.000	3667.772	0.000	0.979	0.000	1752.509	0.000
H	0.989	0.000	254.399	0.000	0.991	0.000	236.175	0.000	0.993	0.000	207.465	0.000
SEX												
M	0.987	0.000	5.122.707	0.000	0.986	0.000	7.368.572	0.000	0.989	0.000	3.378.040	0.000
F	0.987	0.000	6.026.447	0.000	0.990	0.000	7.677.271	0.000	0.994	0.000	4.523.133	0.000

	ES				LC				HS				
	SHAPIRO-WILK		D'AGOSTINO-PEARSON'S		SHAPIRO-WILK		D'AGOSTINO-PEARSON'S		SHAPIRO-WILK		D'AGOSTINO-PEARSON'S		
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	
SC_TP													
A	0.988	0.000	8019.444	0.000	0.986	0.000	14203.329	0.000	0.991	0.000	7069.660	0.000	
B	0.988	0.000	344.713	0.000	0.980	0.000	1009.799	0.000	0.987	0.000	419.200	0.000	
C	0.988	0.000	201.528	0.000	0.984	0.000	489.678	0.000	0.987	0.000	216.877	0.000	
D	0.992	0.000	285.611	0.000	0.966	0.000	6819.560	0.000	0.969	0.000	4645.034	0.000	
E	0.990	0.000	138.684	0.000	0.981	0.000	818.439	0.000	0.979	0.000	540.138	0.000	
RACE													
A	0.991	0.000	352.002	0.000	0.985	0.000	968.664	0.000	0.991	0.000	757.144	0.000	
B	0.989	0.000	1751.176	0.000	0.988	0.000	2608.077	0.000	0.993	0.000	1577.074	0.000	
C	0.988	0.000	7786.179	0.000	0.988	0.000	12065.974	0.000	0.992	0.000	5521.982	0.000	
D	0.988	0.000	113.730	0.000	0.991	0.000	124.531	0.000	0.991	0.000	128.569	0.000	
E	0.992	0.000	9.031	0.011	0.987	0.000	24.979	0.000	0.990	0.000	18.067	0.000	
F	0.988	0.000	180.864	0.000	0.986	0.000	298.013	0.000	0.991	0.000	142.489	0.000	
AGE													
16	0.992	0.000	386.881	0.000	0.984	0.000	1.602.990	0.000	0.984	0.000	949.981	0.000	
17	0.991	0.000	1.636.296	0.000	0.989	0.000	3.488.876	0.000	0.990	0.000	2.381.776	0.000	
18	0.987	0.000	1.781.937	0.000	0.991	0.000	2.074.967	0.000	0.993	0.000	1.452.235	0.000	
19	0.984	0.000	1.299.520	0.000	0.990	0.000	1.333.308	0.000	0.994	0.000	823.759	0.000	
20	0.982	0.000	1.218.373	0.000	0.990	0.000	892.774	0.000	0.994	0.000	520.399	0.000	
21	0.983	0.000	890.581	0.000	0.990	0.000	712.993	0.000	0.994	0.000	335.733	0.000	
22	0.983	0.000	688.225	0.000	0.989	0.000	597.870	0.000	0.993	0.000	278.814	0.000	
23	0.983	0.000	535.734	0.000	0.988	0.000	483.427	0.000	0.992	0.000	239.909	0.000	
24	0.983	0.000	435.156	0.000	0.987	0.000	423.949	0.000	0.992	0.000	207.531	0.000	
25	0.985	0.000	309.684	0.000	0.988	0.000	338.846	0.000	0.993	0.000	101.155	0.000	

	NS				MA			
	SHAPIRO-WILK		D'AGOSTINO-PEARSON'S		SHAPIRO-WILK		D'AGOSTINO-PEARSON'S	
	Statistics	p-value	Statistics	p-value	Statistics	p-value	Statistics	p-value
ST_FA								
A	0.989	0.000	211.464	0.000	0.954	0.000	650.446	0.000
B	0.991	0.000	1829.843	0.000	0.972	0.000	3354.775	0.000
C	0.991	0.000	1610.361	0.000	0.978	0.000	2525.070	0.000
D	0.992	0.000	1173.432	0.000	0.979	0.000	2206.588	0.000
E	0.993	0.000	2668.791	0.000	0.984	0.000	3519.617	0.000
F	0.991	0.000	797.031	0.000	0.992	0.000	1213.530	0.000
G	0.989	0.000	516.013	0.000	0.992	0.000	897.848	0.000
H	0.990	0.000	629.238	0.000	0.971	0.000	1496.660	0.000
ST_MO								
A	0.989	0.000	242.305	0.000	0.954	0.000	549.018	0.000
B	0.990	0.000	1693.101	0.000	0.970	0.000	2750.805	0.000
C	0.991	0.000	1637.741	0.000	0.977	0.000	2505.696	0.000
D	0.991	0.000	1350.691	0.000	0.978	0.000	2421.978	0.000
E	0.992	0.000	2839.977	0.000	0.983	0.000	4078.049	0.000
F	0.992	0.000	893.707	0.000	0.990	0.000	1433.383	0.000
G	0.991	0.000	939.956	0.000	0.991	0.000	1547.341	0.000
H	0.988	0.000	208.637	0.000	0.960	0.000	706.511	0.000
SEX								
M	0.994	0.000	1.975.530	0.000	0.986	0.000	4.717.196	0.000
F	0.989	0.000	6.036.796	0.000	0.970	0.000	13.942.656	0.000

	NS				MA			
	SHAPIRO-WILK		D'AGOSTINO-PEARSON'S		SHAPIRO-WILK		D'AGOSTINO-PEARSON'S	
	Statistics	p-value	Statistics	p-value	Statistics	p-value	Statistics	p-value
SC_TP								
A	0.991	0.000	6968.476	0.000	0.977	0.000	12782.299	0.000
B	0.993	0.000	371.909	0.000	0.982	0.000	571.375	0.000
C	0.992	0.000	167.959	0.000	0.983	0.000	238.304	0.000
D	0.986	0.000	1627.098	0.000	0.994	0.000	1326.280	0.000
E	0.994	0.000	88.956	0.000	0.993	0.000	205.623	0.000
RACE								
A	0.988	0.000	745.230	0.000	0.965	0.000	1305.368	0.000
B	0.989	0.000	1631.412	0.000	0.970	0.000	3791.862	0.000
C	0.993	0.000	5704.618	0.000	0.979	0.000	11365.635	0.000
D	0.989	0.000	177.351	0.000	0.976	0.000	229.642	0.000
E	0.990	0.000	9.516	0.009	0.967	0.000	50.613	0.000
F	0.990	0.000	294.492	0.000	0.974	0.000	357.758	0.000
AGE								
16	0.990	0.000	1.243.691	0.000	0.987	0.000	862.187	0.000
17	0.991	0.000	2.907.707	0.000	0.979	0.000	3.754.267	0.000
18	0.989	0.000	1.923.554	0.000	0.973	0.000	3.654.484	0.000
19	0.987	0.000	1.226.818	0.000	0.969	0.000	2.281.556	0.000
20	0.986	0.000	816.545	0.000	0.971	0.000	1.589.416	0.000
21	0.988	0.000	527.946	0.000	0.972	0.000	1.027.080	0.000
22	0.991	0.000	309.016	0.000	0.975	0.000	715.323	0.000
23	0.992	0.000	227.488	0.000	0.976	0.000	544.272	0.000
24	0.992	0.000	151.151	0.000	0.975	0.000	439.712	0.000
25	0.993	0.000	132.403	0.000	0.977	0.000	319.462	0.000

APPENDIX C – MANN-WHITNEY TEST RESULTS

		ST_FA												
ES	B		C		D		E		F		G		H	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
A	-18.20	0.000	-30.83	0.000	-36.17	0.000	-51.45	0.000	-72.41	0.000	-79.18	0.000	-21.89	0.000
B			-30.03	0.000	-42.11	0.000	-87.39	0.000	-116.69	0.000	-120.44	0.000	-9.65	0.000
C					-13.41	0.000	-53.73	0.000	-93.94	0.000	-101.83	0.000	-13.16	0.000
D							-36.16	0.000	-80.39	0.000	-90.37	0.000	-23.14	0.000
E									-60.54	0.000	-74.19	0.000	-52.77	0.000
F											-19.91	0.000	-87.13	0.000
G													-95.56	0.000
LC	B		C		D		E		F		G		H	
A	-11.80	0.000	-20.65	0.000	-27.56	0.000	-48.13	0.000	-76.32	0.000	-84.03	0.000	-21.60	0.000
B			-20.53	0.000	-35.89	0.000	-93.74	0.000	-134.45	0.000	-137.52	0.000	-20.40	0.000
C					-16.17	0.000	-70.59	0.000	-119.44	0.000	-125.69	0.000	-4.90	0.000
D							-49.47	0.000	-104.09	0.000	-113.28	0.000	-7.69	0.000
E									-77.47	0.000	-92.20	0.000	-45.81	0.000
F											-24.02	0.000	-93.18	0.000
G													-103.24	0.000
HS	B		C		D		E		F		G		H	
A	-8.53	0.000	-16.72	0.000	-21.46	0.000	-39.94	0.000	-70.52	0.000	-80.02	0.000	-13.77	0.000
B			-18.81	0.000	-29.07	0.000	-79.93	0.000	-126.56	0.000	-133.05	0.000	-11.01	0.000
C					-10.96	0.000	-58.49	0.000	-112.21	0.000	-121.64	0.000	-3.27	0.001
D							-43.61	0.000	-100.72	0.000	-112.23	0.000	-11.69	0.000
E									-77.78	0.000	-94.20	0.000	-45.87	0.000
F											-25.50	0.000	-94.13	0.000
G													-105.59	0.000

ST_MO

LC	B		C		D		E		F		G		H	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
A	-11.16	0.000	-18.17	0.000	-23.01	0.000	-42.67	0.000	-68.93	0.000	-71.68	0.000	-8.07	0.000
B			-15.75	0.000	-26.76	0.000	-79.95	0.000	-123.57	0.000	-130.46	0.000	-0.77	0.220
C					-11.32	0.000	-64.66	0.000	-113.81	0.000	-121.16	0.000	-8.67	0.000
D							-51.87	0.000	-105.13	0.000	-112.70	0.000	-14.16	0.000
E									-76.69	0.000	-86.09	0.000	-36.76	0.000
F											-7.90	0.000	-67.55	0.000
G													-70.97	0.000
HS														
	B		C		D		E		F		G		H	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
A	-8.09	0.000	-14.06	0.000	-16.16	0.000	-34.46	0.000	-62.97	0.000	-68.09	0.000	-5.68	0.000
B			-13.28	0.000	-18.01	0.000	-65.88	0.000	-114.86	0.000	-126.43	0.000	-0.92	0.179
C					-4.89	0.000	-52.81	0.000	-106.64	0.000	-118.79	0.000	-7.66	0.000
D							-47.16	0.000	-102.64	0.000	-114.87	0.000	-10.06	0.000
E									-77.53	0.000	-92.16	0.000	-31.05	0.000
F											-12.02	0.000	-63.92	0.000
G													-69.97	0.000
NS														
	B		C		D		E		F		G		H	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
A	-6.67	0.000	-12.52	0.000	-15.02	0.000	-33.47	0.000	-63.84	0.000	-70.06	0.000	-6.67	0.000
B			-12.96	0.000	-18.60	0.000	-66.57	0.000	-118.37	0.000	-132.06	0.000	-2.20	0.014
C					-5.90	0.000	-53.93	0.000	-110.37	0.000	-124.66	0.000	-4.29	0.000
D							-46.82	0.000	-104.86	0.000	-119.18	0.000	-7.19	0.000
E									-79.87	0.000	-96.76	0.000	-27.98	0.000
F											-13.76	0.000	-62.72	0.000
G													-69.97	0.000

ST_MO

MA	B		C		D		E		F		G		H	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
A	-20.50	0.000	-32.63	0.000	-36.71	0.000	-54.42	0.000	-77.55	0.000	-81.51	0.000	-19.61	0.000
B			-27.32	0.000	-36.87	0.000	-87.55	0.000	-127.66	0.000	-137.53	0.000	-5.31	0.000
C					-9.98	0.000	-60.08	0.000	-110.53	0.000	-121.21	0.000	-8.55	0.000
D							-48.80	0.000	-102.85	0.000	-113.68	0.000	-13.43	0.000
E									-76.55	0.000	-89.39	0.000	-34.81	0.000
F											-10.35	0.000	-66.00	0.000
G													-71.25	0.000

SEX	
	F
ES	
M	Statistic -81.02 p-value 0.000
LC	
M	Statistic -4.27 p-value 0.000
HS	
M	Statistic -72.51 p-value 0.000
NS	
M	Statistic -81.38 p-value 0.000
MA	
M	Statistic -129.49 p-value 0.000

SC_TP								
ES	B		C		D		E	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
A	-26.79	0.000	-15.37	0.000	-185.09	0.000	-64.49	0.000
B			-4.19	0.000	-70.49	0.000	-31.33	0.000
C					-61.18	0.000	-31.92	0.000
D							-27.53	0.000
LC	B		C		D		E	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
A	-47.43	0.000	-19.90	0.000	-208.39	0.000	-64.61	0.000
B			-13.82	0.000	-67.55	0.000	-17.90	0.000
C					-69.06	0.000	-29.35	0.000
D							-40.94	0.000
HS	B		C		D		E	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
A	-43.91	0.000	-17.31	0.000	-206.74	0.000	-65.05	0.000
B			-13.48	0.000	-68.60	0.000	-20.13	0.000
C					-69.65	0.000	-31.10	0.000
D							-39.91	0.000
NS	B		C		D		E	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
A	-48.02	0.000	-20.73	0.000	-223.81	0.000	-69.23	0.000
B			-13.08	0.000	-76.36	0.000	-21.58	0.000
C					-74.24	0.000	-31.47	0.000
D							-44.81	0.000

SC_TP								
MA	B		C		D		E	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
A	-36.45	0.000	-14.34	0.000	-208.60	0.000	-62.95	0.000
B			-10.91	0.000	-76.82	0.000	-24.32	0.000
C					-72.33	0.000	-31.77	0.000
D							-41.77	0.000

RACE												
ES	B		C		D		E		F			
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
A	-12.41	0.000	-50.57	0.000	-22.76	0.000	-5.16	0.000	-20.34	0.000		
B			-61.99	0.000	-18.05	0.000	-7.71	0.000	-14.54	0.000		
C					-0.78	0.219	-14.41	0.000	-7.25	0.000		
D							-13.39	0.000	-5.11	0.000		
E									-11.66	0.000		
LC	B		C		D		E		F			
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
A	-10.77	0.000	-53.07	0.000	-25.49	0.000	-4.70	0.000	-32.94	0.000		
B			-69.17	0.000	-21.81	0.000	-6.95	0.000	-29.82	0.000		
C					-2.62	0.004	-14.55	0.000	-7.18	0.000		
D							-14.12	0.000	-2.44	0.007		
E									-15.32	0.000		

RACE											
HS	B		C		D		E		F		
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	
A	-15.36	0.000	-55.75	0.000	-28.11	0.000	-2.17	0.015	-35.83	0.000	
B			-65.70	0.000	-22.49	0.000	-5.48	0.000	-30.37	0.000	
C					-4.49	0.000	-12.99	0.000	-8.81	0.000	
D							-13.26	0.000	-2.01	0.022	
E									-14.52	0.000	
NS											
	B		C		D		E		F		
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	
A	-17.55	0.000	-61.75	0.000	-32.41	0.000	-0.33	0.370	-35.87	0.000	
B			-71.43	0.000	-26.15	0.000	-3.45	0.000	-28.81	0.000	
C					-7.08	0.000	-11.73	0.000	-4.80	0.000	
D							-12.82	0.000	-2.63	0.004	
E									-12.34	0.000	
MA											
	B		C		D		E		F		
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	
A	-25.98	0.000	-73.93	0.000	-39.00	0.000	-1.32	0.094	-37.80	0.000	
B			-78.16	0.000	-29.50	0.000	-4.24	0.000	-26.08	0.000	
C					-8.43	0.000	-13.14	0.000	-0.50	0.309	
D							-14.66	0.000	-6.76	0.000	
E									-12.09	0.000	

		AGE									
ES		17	18	19	20	21	22	23	24	25	
		Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value
16		-2.47 0.007	-5.77 0.000	-10.61 0.000	-19.26 0.000	-24.15 0.000	-28.36 0.000	-29.59 0.000	-28.68 0.000	-25.98 0.000	
17			-11.06 0.000	-16.08 0.000	-25.63 0.000	-30.14 0.000	-33.93 0.000	-34.33 0.000	-32.50 0.000	-29.02 0.000	
18				-6.44 0.000	-16.53 0.000	-21.76 0.000	-26.17 0.000	-27.32 0.000	-26.20 0.000	-23.39 0.000	
19					-9.66 0.000	-15.13 0.000	-19.85 0.000	-21.59 0.000	-21.09 0.000	-18.90 0.000	
20						-5.86 0.000	-11.03 0.000	-13.48 0.000	-13.71 0.000	-12.22 0.000	
21							-5.30 0.000	-8.11 0.000	-8.80 0.000	-7.75 0.000	
22								-3.05 0.001	-4.08 0.000	-3.39 0.000	
23									-1.17 0.120	-0.69 0.246	
24										-0.39 0.347	
LC											
		Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value
16		-13.87 0.000	-23.50 0.000	-16.30 0.000	-13.66 0.000	-12.10 0.000	-9.26 0.000	-8.37 0.000	-4.28 0.000	-3.11 0.001	
17			-14.34 0.000	-5.74 0.000	-3.23 0.001	-2.24 0.013	-0.24 0.404	-0.43 0.334	-4.12 0.000	-4.67 0.000	
18				-5.83 0.000	-7.01 0.000	-6.87 0.000	-8.43 0.000	-7.81 0.000	-10.70 0.000	-10.68 0.000	
19					-1.61 0.054	-2.01 0.022	-3.88 0.000	-3.73 0.000	-6.86 0.000	-7.18 0.000	
20						-0.52 0.303	-2.41 0.008	-2.40 0.008	-5.57 0.000	-5.98 0.000	
21							-1.84 0.033	-1.88 0.030	-4.98 0.000	-5.44 0.000	
22								-0.16 0.436	-3.26 0.001	-3.81 0.000	
23									-2.96 0.002	-3.51 0.000	
24										-0.70 0.243	
HS											
		Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value
16		-6.86 0.000	-21.63 0.000	-19.10 0.000	-20.49 0.000	-20.25 0.000	-17.97 0.000	-15.93 0.000	-10.97 0.000	-8.17 0.000	
17			-20.60 0.000	-16.78 0.000	-18.19 0.000	-17.63 0.000	-14.98 0.000	-12.80 0.000	-7.60 0.000	-4.76 0.000	
18				-0.21 0.416	-3.01 0.001	-4.07 0.000	-2.76 0.003	-1.87 0.031	-2.08 0.019	-4.08 0.000	
19					-2.85 0.002	-3.85 0.000	-2.69 0.004	-1.88 0.030	-1.83 0.033	-3.79 0.000	
20						-1.16 0.124	-0.25 0.402	-0.32 0.375	-3.71 0.000	-5.47 0.000	
21							-0.79 0.214	-1.28 0.101	-4.48 0.000	-6.15 0.000	
22								-0.52 0.303	-3.66 0.000	-5.33 0.000	
23									-3.03 0.001	-4.65 0.000	
24										-1.70 0.045	

		AGE									
NS		17	18	19	20	21	22	23	24	25	
		Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value
16		-12.78	-24.33	-21.80	-22.58	-22.26	-20.55	-18.08	-15.64	-11.89	0.000
17			-16.58	-13.77	-14.96	-14.81	-13.01	-10.72	-8.61	-5.13	0.000
18				-0.04	-2.63	-3.77	-2.87	0.056	-0.43	0.333	0.013
19					-2.31	-3.38	-2.58	0.079	-0.35	0.362	0.015
20						-1.19	0.117	-0.43	0.334	-1.30	0.000
21							-0.55	0.290	-2.20	0.014	0.000
22								-0.88	0.188	-1.71	0.044
23									-0.84	0.201	0.001
24										-2.16	0.015
MA											
		Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value	Statistic p-value
16		-14.93	-27.87	-27.46	-25.51	-25.67	-24.32	-23.09	-22.16	-17.81	0.000
17			-18.89	-18.93	-16.79	-17.51	-16.26	-15.31	-14.89	-10.77	0.000
18				-3.10	-2.80	-5.17	-5.05	-5.18	-5.83	-2.67	0.004
19					-0.03	0.489	-2.64	-2.97	-3.81	-1.01	0.157
20						-2.33	-2.47	-2.83	-3.66	-0.95	0.171
21							-0.30	0.214	-1.76	0.039	0.000
22								-0.79	0.214	-0.69	0.244
23								-0.50	0.309	-0.91	0.181
24									-0.95	0.171	-1.30
										-2.07	0.019

APPENDIX D – sMC AND VIP VALUES

Variable	ES		LC		HS		NS		MA	
	sMC	VIP	sMC	VIP	sMC	VIP	sMC	VIP	sMC	VIP
AGE	0.021	-	0.002	-	0.001	-	0.000	-	0.006	-
SEX_M	5707.661	1.664	503.227	0.046	925.951	1.271	1707.030	1.434	10437.647	2.324
RACE_B	1047.005	0.878	1344.721	0.889	1632.630	0.803	1275.279	0.861	3030.721	0.961
RACE_C	710.864	1.323	1058.677	1.211	1285.917	1.137	1200.583	1.209	3105.124	1.382
RACE_D	11593.757	0.025	16466.935	0.035	19703.270	0.034	36561.891	0.051	98264.323	0.064
RACE_E	21564.101	0.024	17058.889	0.024	9457.157	0.019	2381.204	0.016	684.578	0.016
RACE_F	25009.480	0.018	70993.796	0.079	74163.562	0.079	57654.616	0.060	80455.299	0.041
LG_E	1383.472	2.235	4184.511	3.047	3729.836	2.857	1891.869	2.490	2224.961	2.702
SC_TP_B	1432.920	0.065	5083.937	0.197	3781.141	0.166	11159.206	0.174	38.052	0.098
SC_TP_C	24589.357	0.006	18733.479	0.026	16156.491	0.010	24847.770	0.013	4811.679	0.024
SC_TP_D	61531.910	2.719	42392.906	2.645	50224.653	2.545	96333.155	2.775	68178.287	2.733
SC_TP_E	131000.270	0.349	94830.983	0.292	107824.185	0.289	120678.728	0.299	109020.004	0.263
HS_ST_B	14069.296	0.102	26986.878	0.142	22023.232	0.084	18324.977	0.013	14877.479	0.138
HS_ST_C	12.547	0.152	4527.229	1.046	2786.064	0.646	2587.247	0.632	1608.477	0.312
HS_ST_D	7894.582	0.632	1603.787	0.133	988.745	0.077	423.965	0.131	88.804	0.149
HS_ST_E	10014.609	0.238	123.105	0.117	135.837	0.048	13.123	0.068	375.414	0.073
HS_ST_F	7257.196	0.094	0.553	0.121	260.977	0.027	1.107	0.059	1209.375	0.081
HS_ST_G	7923.490	0.003	367.876	0.107	0.047	0.039	312.924	0.046	1272.146	0.033
HS_ST_H	4577.567	0.063	4479.041	0.091	581.180	0.025	1267.683	0.023	2689.300	0.007
HS_ST_I	11542.467	0.047	25166.059	0.094	16059.173	0.053	13124.350	0.038	17066.190	0.018
HS_ST_J	6644.407	0.073	25988.196	0.052	16351.714	0.019	6949.886	0.004	14084.757	0.015
HS_ST_K	5206.708	0.063	44421.116	0.071	33958.897	0.044	27003.001	0.031	24311.970	0.002
HS_ST_L	286.582	0.060	46484.250	0.056	44322.665	0.040	37369.065	0.032	30377.192	0.002
HS_ST_M	1764.209	0.065	65084.969	0.058	61784.685	0.045	45627.129	0.032	36519.176	0.003
HS_ST_N	390.878	0.510	15976.883	0.362	19218.037	0.377	8728.301	0.221	2016.293	0.110
IN_B	177.836	0.561	0.112	-	31.966	0.588	58.123	0.571	67.034	0.619
IN_C	482.010	0.316	490.964	0.233	542.780	0.258	108.063	0.335	725.413	0.279
IN_D	3453.209	0.004	2206.716	0.053	2571.879	0.027	1490.679	0.015	4206.395	0.041
IN_E	5806.535	0.272	3546.197	0.313	4403.106	0.293	3300.978	0.262	5657.622	0.283
IN_F	9639.088	0.365	5303.903	0.391	7663.117	0.390	5586.333	0.371	10830.755	0.400

Variable	ES		LC		HS		NS		MA	
	sMC	VIP	sMC	VIP	sMC	VIP	sMC	VIP	sMC	VIP
IN_G	26517.171	0.466	15301.390	0.468	19204.703	0.462	16440.650	0.471	29878.199	0.499
IN_H	46621.200	0.372	27893.967	0.367	30305.388	0.356	30042.985	0.376	50760.857	0.384
IN_I	84211.711	0.284	53415.224	0.276	66436.049	0.277	60112.693	0.281	93557.880	0.290
IN_J	86719.069	0.244	51373.931	0.230	61927.267	0.229	59239.973	0.240	98250.563	0.250
IN_K	81037.432	0.211	51864.272	0.205	57831.113	0.203	47766.434	0.209	98312.456	0.224
IN_L	72142.375	0.335	55307.635	0.338	57024.343	0.327	39170.403	0.330	80772.648	0.344
IN_M	89471.452	0.258	60194.232	0.251	74147.951	0.251	60496.160	0.257	107846.629	0.272
IN_N	93864.827	0.218	87224.119	0.225	92144.538	0.221	89390.513	0.233	123932.348	0.243
IN_O	98471.260	0.322	84698.719	0.325	80476.427	0.318	74485.172	0.336	128158.089	0.370
ST_FA_B	1.151	1.389	30.253	1.280	25.981	1.093	107.988	1.092	1.498	-
ST_FA_C	463.356	0.678	1.202	-	3.291	-	3.619	0.714	716.381	0.683
ST_FA_D	1032.656	0.245	199.501	0.416	97.712	0.390	1.604	-	945.613	0.350
ST_FA_E	1197.786	0.832	778.059	0.839	343.429	0.597	196.894	0.510	1507.711	0.700
ST_FA_F	9011.570	1.132	10881.417	1.186	8302.974	1.121	5956.667	1.126	16717.112	1.155
ST_FA_G	13199.381	1.045	20088.187	1.062	20370.797	1.033	19044.664	1.097	42846.770	1.132
ST_FA_H	159.514	0.456	653.667	0.354	3.526	-	15.857	0.398	1028.480	0.426
ST_MO_B	26.475	1.226	9.946	1.107	0.113	-	46.261	0.944	0.487	-
ST_MO_C	1268.739	0.765	480.898	0.863	577.300	0.767	87.164	0.784	1773.949	0.801
ST_MO_D	1273.469	0.574	534.389	0.690	641.018	0.686	130.287	0.685	2214.518	0.632
ST_MO_E	1786.032	0.502	1240.794	0.545	1241.614	0.297	529.467	0.216	3515.436	0.406
ST_MO_F	12593.689	1.146	9861.313	1.193	12265.066	1.109	6844.746	1.140	26149.027	1.198
ST_MO_G	10609.565	1.366	5320.634	1.294	10132.207	1.285	5000.721	1.339	19670.720	1.380
ST_MO_H	4102.886	0.206	8322.719	0.184	495.254	0.154	3599.269	0.136	604.665	0.157
OC_FA_B	4.448	-	1034.147	0.658	170.316	0.811	400.669	0.870	447.698	0.861
OC_FA_C	4.568	0.915	1237.831	0.761	6.618	-	32.133	1.017	1.674	-
OC_FA_D	1968.136	1.879	5879.808	1.957	720.202	1.743	547.822	1.729	476.911	1.748
OC_FA_E	11204.561	1.295	17916.088	1.291	1390.474	1.244	1844.547	1.314	1020.781	1.376
OC_FA_F	543.294	0.511	480.806	0.412	930.318	0.452	868.136	0.449	1635.239	0.477
OC_MO_B	33.613	1.606	0.142	-	397.502	1.769	659.581	1.920	308.440	1.888
OC_MO_C	34.513	0.271	2.496	0.244	574.074	0.278	1374.478	0.333	119.449	0.210

Variable	ES		LC		HS		NS		MA	
	sMC	VIP	sMC	VIP	sMC	VIP	sMC	VIP	sMC	VIP
OC_MO_B	33.613	1.606	0.142	-	397.502	1.769	659.581	1.920	308.440	1.888
OC_MO_C	34.513	0.271	2.496	0.244	574.074	0.278	1374.478	0.333	119.449	0.210
OC_MO_D	55.316	2.067	239.024	2.077	141.437	1.920	160.343	1.929	90.253	1.959
OC_MO_E	4887.345	0.876	5422.684	0.864	235.465	0.834	206.104	0.884	1034.041	0.938
OC_MO_F	918.173	0.268	344.520	0.172	2268.422	0.154	2025.276	0.108	1704.792	0.150
RES_B	136.352	0.154	125.977	0.203	368.324	0.022	82.215	0.040	252.366	0.140
RES_C	243.045	0.185	379.392	0.176	409.870	0.144	199.451	0.156	391.596	0.116
RES_D	541.576	0.604	798.395	0.210	778.683	0.310	394.415	0.382	715.099	0.475
RES_E	3043.909	0.211	2846.378	0.330	2656.288	0.241	1916.636	0.229	1839.216	0.152
RES_F	12362.599	0.189	10717.611	0.218	10595.147	0.200	7986.501	0.201	7567.294	0.168
RES_G	26102.611	0.079	32949.605	0.101	26875.825	0.089	32463.848	0.098	31316.037	0.087
RES_H	49570.125	0.040	16475.540	0.033	37154.851	0.039	29484.786	0.038	49278.447	0.038
RES_I	34538.145	0.016	22210.365	0.016	10688.468	0.012	28809.450	0.017	25621.635	0.013
RES_J	16469.612	0.012	27995.284	0.018	21079.811	0.014	19525.325	0.014	15995.851	0.012
CAR_Y	51.440	1.534	137.435	1.167	84.641	1.372	5.759	-	1.709	-
MOTO_Y	782.693	0.624	897.353	0.764	483.053	0.494	118.897	0.351	170.984	0.271
CELL_B	144.531	1.243	385.733	1.128	207.617	0.932	42.578	0.800	229.412	1.056
PC_Y	876.295	2.050	508.545	2.092	873.232	2.103	592.293	1.962	445.563	2.006
INT_Y	277.776	1.667	8.799	1.548	29.013	1.476	5.423	1.359	0.405	-
CB_TV_Y	0.093	-	3.909	-	30.433	1.894	105.228	1.842	27.854	2.007
TV_B	0.036	-	19.402	2.147	8.412	1.862	17.001	1.846	3.570	-
DVD_Y	0.980	-	0.098	-	4.031	-	2.143	-	0.114	-
TP_Y	1.436	2.071	109.416	2.385	20.586	2.119	24.231	2.099	2.541	-
HK_Y	64.945	1.182	133.760	1.097	2.619	-	74.879	1.204	709.148	1.292
BATH_B	153.060	2.790	38.344	2.585	98.219	2.582	179.439	2.714	179.179	2.877
BED_B	6.046	-	41.084	0.171	0.029	-	1.420	-	49.652	0.373
MW_Y	6.584	-	3.343	-	49.486	1.321	32.603	1.298	0.017	-
FGE_B	35.803	0.526	142.392	0.469	0.610	-	1.901	-	63.827	0.596
FZR_Y	614.163	2.169	346.997	1.626	372.921	1.645	238.805	1.601	649.861	1.978
WA_MA_Y	26.885	0.946	16.848	0.894	14.325	0.843	3.445	-	4.656	-
DR_MA_Y	3.769	-	2.560	-	0.885	-	19.411	1.171	1.217	-
DI_WA_Y	899.557	0.723	455.690	0.727	139.645	0.701	1370.592	0.769	1289.074	0.796
VA_CL_Y	15.740	2.350	100.568	2.492	161.668	2.395	196.229	2.436	188.340	2.597

