UNIVERSIDADE FEDERAL DO PARANÁ

SETOR DE CIÊNCIAS SOCIAIS

DEPARTAMENTO DE CIÊNCIAS ECONÔMICAS

**ECONOMETRIA – SE308 – Prof. Flávio de Oliveira Gonçalves**

**LETÍCIA RODRIGUES FERREIRA SENGUE – GRR20195213**

5º Relatório Final – Econometria

1. **INTRODUÇÃO**

O Enade (Exame Nacional de Desempenho de Estudantes) é uma avaliação realizada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), órgão vinculado ao Ministério da Educação (MEC) do Brasil. O ENADE foi instituído em 2004 como parte do Sistema Nacional de Avaliação da Educação Superior (SINAES), criado pela Lei nº 10.861/2004. Ele se tornou obrigatório para os estudantes ingressantes e concluintes dos cursos de graduação, abrangendo diversas áreas do conhecimento, com o objetivo de avaliar o desempenho dos estudantes em relação aos conteúdos programáticos, habilidades e competências desenvolvidas ao longo dos cursos de graduação.

Além disso a função do Enade é utilizar as informações obtidas para promover a melhoria da qualidade da educação superior como um todo e na formação dos estudantes. Dado o valor da avaliação, a falta de motivação dos participantes poderia comprometer a precisão dos resultados do exame, o que teria impactos negativos tanto nas instituições de ensino quanto no próprio Sinaes.

O objetivo geral deste relatório é utilizar modelos estatísticos para avaliar se há correlação entre a abstenção e a entrega de provas em branco pelos estudantes de economia no ENADE 2018, com a insatisfação desses estudantes com a qualidade do suporte oferecido pela universidade ao longo da graduação.

1. **METODOLOGIA**

Para alcançar os objetivos estabelecidos na introdução, a base de dados foi construída utilizando as informações dos microdados fornecidos pelo Enade, que incluem as respostas dos questionários preenchidos pelos alunos. A partir disso, serão analisados os seguintes identificadores da base de dados:

* Avaliações de aprendizagem realizadas durante o curso foram compatíveis com os conteúdos ou temas trabalhados pelos professores?
* A coordenação do curso esteve disponível para orientação acadêmica dos estudantes?

Portanto, por meio de um estimador de máxima verossimilhança faremos a estimação dos modelos Logit e Probit, a fim de obter as respectivas razões de chance e probabilidades relacionadas a hipótese proposta na introdução, onde a variável dependente é: tipo de presença na prova. A análise das regressões, assim como as métricas inerentes a elas, foram realizadas com suporte do software estatístico Rstudio que permite o gerenciamento, manipulação e visualização dos dados buscando identificar variáveis estatisticamente significativas que impactam na adesão dos estudantes à prova do ENADE. Anteriormente à análise dos dados neste programa, foi necessária a tabulação dos dados com o auxílio da ferramenta Microsoft Excel, onde foi efetuada a conversão de variáveis qualitativas em variáveis numéricas binárias (dummies). Para este estudo, foi efetuada a codificação de algumas variáveis obtidas na base de dados utilizada nesta pesquisa, convertendo a resposta da variável para valores binários atribuindo valores 0 (zero) e 1 (um) para as variáveis relacionadas, conforme descrito abaixo.

Variável dependente:

|  |  |  |
| --- | --- | --- |
| Tipo de presença na prova | 222 = Prova não realizada devido a ausência do estudante | 1 |
| 333 = Participação com prova em branco |
| 555 = Participação com respostas na prova | 0 |

Variáveis independentes:

|  |  |  |
| --- | --- | --- |
| **PERGUNTA** | **RESPOSTA** | **DUMMY** |
| As avaliações de aprendizagem realizadas durante o curso foram compatíveis com os conteúdos ou temas trabalhados pelos professores. | 1 = Discordo totalmente. | 0 |
| 2 = Discordo. |
| 3 = Discordo parcialmente. |
| 4 = Concordo parcialmente. | 1 |
| 5 = Concordo. |
| 6 = Concordo totalmente. |
| A coordenação do curso esteve disponível para orientação acadêmica dos estudantes. | 1 = Discordo totalmente. | 0 |
| 2 = Discordo. |
| 3 = Discordo parcialmente. |
| 4 = Concordo parcialmente. | 1 |
| 5 = Concordo. |
| 6 = Concordo totalmente. |

Com o objetivo de incluir respostas que agregassem valor à regressão e respondessem à pergunta do questionário, foram excluídas as opções das variáveis independentes "não se aplica" e "não sei responder" e da variável dependente “participação com resultado desconsiderado pela Aplicadora”.

1. **MODELOS**

No modelo Logit, há a transformação logarítmica das chances (razões de chances) do evento ocorrer. Ele assume uma relação linear entre o logaritmo das chances (logits) e as variáveis independentes. Essa diferença é fundamental para modelar efetivamente a relação entre as variáveis e as probabilidades de sucesso.

Enquanto as probabilidades variam de 0 a 1, os logits (valores logarítmicos das chances) não possuem limites específicos. Isso significa que os logits podem variar de -∞ a +∞ à medida que a variável explanatória muda de -∞ a +∞. Além disso, embora o modelo Logit seja linear em relação aos coeficientes das variáveis independentes, as probabilidades em si não são lineares. Isso difere do Modelo de Probabilidade Linear (MPL), onde as probabilidades aumentam linearmente com as variáveis independentes.

O modelo é formulado por:

Uma imagem contendo objeto, relógio

Descrição gerada automaticamente

Onde,



No modelo Probit, a variável dependente é modelada como uma função de distribuição de probabilidade normal. Os coeficientes estimados no modelo Probit têm uma interpretação similar ao modelo Logit. Eles representam a mudança nas chances de sucesso associadas a uma alteração unitária nas variáveis independentes, mantendo as outras variáveis constantes.

Modelo Probit é formulado por:

Uma imagem contendo Texto

Descrição gerada automaticamente

Logit e probit oferecessem, do ponto de vista qualitativo, resultados semelhantes. Embora os modelos sejam semelhantes, deve-se ficar atento ao interpretar os coeficientes estimados pelos dois modelos. Isso porque, embora a distribuição logística padrão (a base do logit) e a normal padrão (a base do probit) tenham ambas média zero, suas variâncias são diferentes; 1 para a normal padrão e pi^2=3 para a logística. No gráfico abaixo é possível visualizar que o Modelo Probit exibe uma curva mais suave e simétrica em relação ao ponto médio, enquanto o Modelo Logit terá uma curva mais inclinada no centro e mais suave nas extremidades:

Diagrama

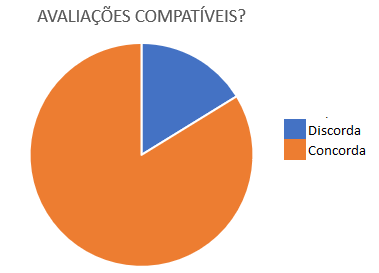
Descrição gerada automaticamente

1. **DESENVOLVIMENTO**

Inicialmente, foi realizada uma análise descritiva dos dados, examinando a frequência e a média das variáveis.

|  |  |  |
| --- | --- | --- |
|  | Coordenação Disponível | Avaliações Compatíveis |
| Concorda | 6240 | 6839 |
| Discorda | 1923 | 1324 |

Gráfico, Gráfico de pizza

Descrição gerada automaticamente 

Médias das variáveis:

Texto

Descrição gerada automaticamente

Analisando as médias, observa-se que a variável “a coordenação do curso esteve disponível para orientação acadêmica dos estudantes” apresenta maior descontentamento em relação a variável “as avaliações de aprendizagem realizadas durante o curso foram compatíveis com os conteúdos ou temas trabalhados pelos professores”.

A base de dados em dummy montada no excel foi importada no software Rstudio, assim, através das regressões do pacote glm, obtive os seguintes resultados para logit e probit:

**LOGIT**:

Texto, Carta

Descrição gerada automaticamente

Texto

Descrição gerada automaticamente

Resultados logit:

Intercepto: O coeficiente estimado para o intercepto é -2.4175, e é altamente significativo (p-valor < 2e-16). Isso indica que, quando todas as outras variáveis independentes são mantidas constantes, a log odds de um estudante não realizar a prova devido à ausência ou participar com prova em branco é muito baixa. Para obter uma interpretação mais intuitiva o intercepto de log odds foi transformado para probabilidades usando a função inversa de logit, que retorna a probabilidade estimada em vez da log odds. No modelo proposto, o exponencial calculado foi de 0,089, ou seja, o odds ratio é igual a 0.089. Sendo odds ratio a razão entre as chances de sucesso (presença na prova) e as chances de fracasso (ausência na prova), quando considerado as variáveis independentes binárias “Coordenação” e “Avaliações compatível” representadas por discordo (valor 0) e concordo (valor 1), para cada aumento de 1 nas variáveis (ou seja, mudando de discordo para concordo), mantendo todas as outras variáveis constantes, espera-se que as chances de presença na prova sejam aproximadamente 0.089 vezes maiores. Assim, entende-se que realizar uma avaliação positiva nas variáveis independentes (concordar) está associado a maior probabilidade de adesão a prova do Enade com chances aproximadamente 0.089 vezes maiores.

Coordenação: O coeficiente estimado para a variável "Coordenação" é -0.1212, mas não é estatisticamente significativo (p-valor = 0.32959). Isso indica que não há evidência suficiente para concluir que a disponibilidade da coordenação do curso para orientação acadêmica dos estudantes tem um impacto significativo na probabilidade de um estudante não realizar a prova ou participar com prova em branco.

Avaliações de aprendizagem compatíveis: O coeficiente estimado para a variável "Professores" é -0.3670, e é estatisticamente significativo (p-valor = 0.00662). Isso sugere que as avaliações de aprendizagem realizadas durante o curso serem compatíveis com os conteúdos ou temas trabalhados pelos professores têm um impacto negativo na probabilidade de um estudante não realizar a prova ou participar com prova em branco.

Os efeitos marginais das variáveis independentes de Xj sobre a probabilidade P, ou seja, a derivada de P em relação a Xj, será diferente em cada ponto da regressão:

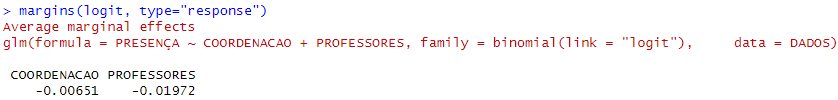
Diagrama

Descrição gerada automaticamente

Para simplificar, foi estimada a média dos efeitos marginais para o conjunto:

Texto

Descrição gerada automaticamente com confiança média



Predicts:

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Inversa de mills:

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente com confiança média

Resumo:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **LOGIT** | REGRESSÃO | EXPONENCIAL | INVERSA DE MILLS | MARGINAL |
| INTERCEPTO | - 2.4175 | 0.0891 | 0.9921 |  |
| COORDENAÇÃO | - 0.1212 | 0.8858 | 0.5482 | -0.00651 |
| AVALIAÇÕES COMPATÍVEIS | - 0.3670 | 0.6928 | 0.6432 | -0.01972 |

**PROBIT:**

Texto, Carta

Descrição gerada automaticamente

Resultados probit:

Intercepto: O coeficiente estimado para o intercepto é -1.39431, é altamente significativo (p-valor < 2e-16). Da mesma forma que no modelo Logit, indica que a probabilidade de um estudante não realizar a prova ou participar com prova em branco é muito baixa, quando as variáveis independentes são mantidas constantes.

Coordenação: O coeficiente estimado para a variável "Coordenação" é -0.05572, mas não é estatisticamente significativo (p-valor = 0.34600). Isso corrobora a conclusão do modelo Logit, indicando que não há evidência suficiente para sugerir que a disponibilidade da coordenação do curso afeta a probabilidade de um estudante não realizar a prova ou participar com prova em branco.

Avaliações de aprendizagem compatíveis: O coeficiente estimado para a variável "Professores" é -0.17663, e é estatisticamente significativo (p-valor = 0.00693). Isso confirma a conclusão do modelo Logit, mostrando que as avaliações de aprendizagem compatíveis com os conteúdos ou temas trabalhados pelos professores têm um impacto negativo na probabilidade de um estudante não realizar a prova ou participar com prova em branco.

Da mesma forma realizada em logit, para simplificar, foi estimada a média dos efeitos marginais para o conjunto:

Uma imagem contendo nome da empresa

Descrição gerada automaticamente

Predicts:

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Inversa de Mills:

Texto

Descrição gerada automaticamente com confiança média

Resumo probit:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PROBIT** | REGRESSÃO | EXPONENCIAL | INVERSA DE MILLS | MARGINAL |
| INTERCEPTO | - 1.3943 |  | 0.9184 |  |
| COORDENAÇÃO | - 0.0557 |  | 0.5222 | -0.00636 |
| AVALIAÇÕES COMPATÍVEIS | - 0.1766 |  | 0.5701 | -0.02016 |

1. **CONCLUSÃO**

Em ambos os modelos a variável “a coordenação do curso esteve disponível para orientação acadêmica dos estudantes” não se mostrou estatisticamente relevante, o que contraria a ideia inicial proposta na introdução, que essa variável seria relevante para inferir que alunos descontentes com a coordenação do curso de graduação estariam mais suscetíveis a entregarem a prova do Enade em branco ou não realizar. Contudo a variável independente “As avaliações de aprendizagem realizadas durante o curso foram compatíveis com os conteúdos ou temas trabalhados pelos professores” mostrou-se estatisticamente relevante nos dois modelos.

No modelo logit a chance do aluno declarar que discorda sobre as avaliações de aprendizagem serem compatíveis com os conteúdos propostos pelos professores é de 36,70% superior para os alunos que entregaram a prova em branco ou estiveram ausentes em relação aos alunos que participaram da prova do Enade.

Em relação ao probit, a chance do aluno declarar que discorda sobre as avaliações de aprendizagem serem compatíveis com os conteúdos propostos pelos professores é de 17,66% superior para os alunos que entregaram a prova em branco ou estiveram ausentes em relação aos alunos que participaram da prova do Enade.

Em relação aos efeitos marginais, a probabilidade de o aluno discordar sobre os métodos de avaliações serem compatíveis é 1,9% para logit 2% para probit maior, quando comparado aos alunos que estiveram presentes na prova do Enade com respostas na prova.

1. **REFERÊNCIAS**

GUJARATI, D.N. e PORTER, D. C. Econometria Básica. 5ª ed. McGraw Hill. Porto Alegre: 2011.