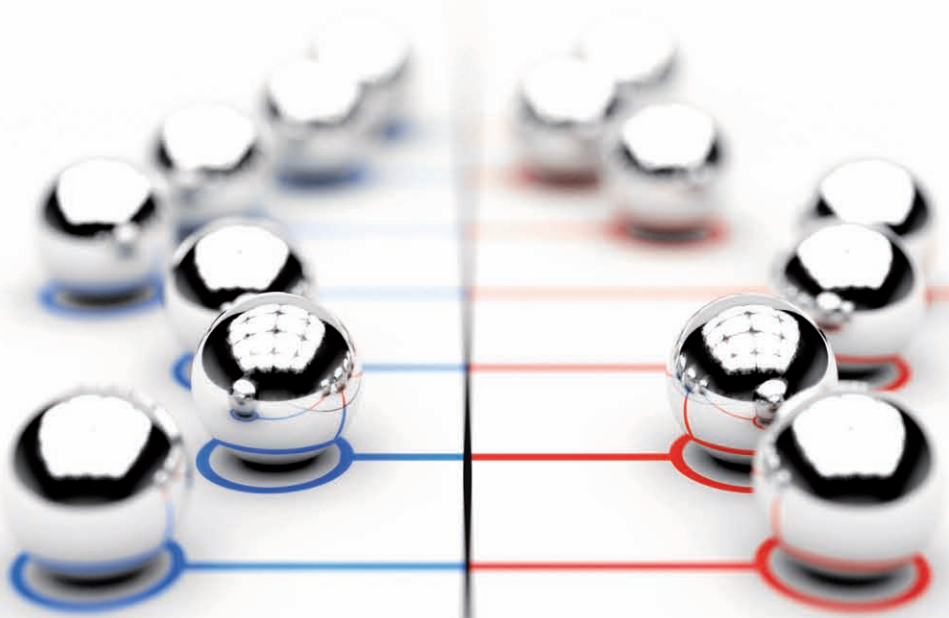


RICARDO ALEXANDRE SARAIVA GOMES

MODELOS ECONOMÉTRICOS DE REGRESSÃO

com variável dependente
qualitativa e limitada

APLICAÇÕES EM SPSS



EDIÇÕES SÍLABO

Modelos econométricos de regressão com variável dependente qualitativa e limitada

Aplicações em SPSS

RICARDO ALEXANDRE SARAIVA GOMES

EDIÇÕES SÍLABO

É expressamente proibido reproduzir, no todo ou em parte, sob qualquer forma ou meio gráfico, eletrónico ou mecânico, inclusive fotocópia, este livro. As transgressões serão passíveis das penalizações previstas na legislação em vigor. Não participe ou encoraje a pirataria eletrónica de materiais protegidos. O seu apoio aos direitos dos autores será apreciado.

Visite a Sílabo na rede

www.silabo.pt

FICHA TÉCNICA:

Título: Modelos Económétricos de Regressão com Variável Dependente Qualitativa e Limitada
– Aplicações em SPSS

Autor: Ricardo Alexandre Saraiva Gomes

© Edições Sílabo, Lda.

Capa: Pedro Mota

1ª Edição – Lisboa, março de 2019

Impressão e acabamentos: ARTIPOL – Artes Tipográficas, Lda.

Depósito Legal: 453640/19

ISBN: 978-989-561-005-1



EDIÇÕES SÍLABO, Lda.

Publicamos conhecimento

Editor: Manuel Robalo

R. Cidade de Manchester, 2

1170-100 Lisboa

Tel.: 218130345

e-mail: silabo@silabo.pt

www.silabo.pt

Índice

Agradecimentos	9
Nota introdutória	11
Capítulo 1	
<hr/>	
O modelo de regressão convencional	13
Capítulo 2	
<hr/>	
O modelo de regressão com variável dependente binária	15
2.1. O modelo Logit ou o modelo Probit	18
2.2. Método de estimação dos coeficientes pelo método da máxima verosimilhança	21
2.3. Ajustamento e comparação de modelos	22
2.4. Teste de Hosmer-Lemeshow	25
2.5. O teste de Wald	25
2.6. Análise de sensibilidade do modelo	26

Capítulo 3

Modelos de regressão com variável dependente limitada e mais de duas categorias	29
3.1. Teste de significância do modelo	31
3.2. Testes à significância dos coeficientes do modelo	32
3.3. Medidas de associação, pseudo R^2	32

Capítulo 4

Aplicação 1 – O caso da aprovação à disciplina de estatística nas licenciaturas	33
4.1. Introdução – enquadramento do estudo	33
4.1.1. Tema	33
4.1.2. Objetivos da aplicação 1	35
4.1.3. Variáveis utilizadas	35
4.1.4. Estatística descritiva de algumas variáveis	37
4.1.5. Metodologia da aplicação 1	42
4.1.6. Hipóteses gerais a considerar em relação aos efeitos principais	44
4.1.7. Hipóteses a considerar em relação à interação de fatores ou variáveis	45
4.2. Resultados obtidos	45
4.2.1. Adequação do modelo em relação aos dados	45
4.2.2. Teste para comparação do modelo final e modelo completo, saturado	48
4.2.3. Análise de resíduos do modelo escolhido e de observações <i>outliers</i>	50
4.2.4. Análise de observações influentes que possam afetar a qualidade do modelo	53
4.2.5. Análise das hipóteses gerais sugeridas	54
4.2.6. Análise da curva ROC	55
4.2.7. Teste de Hosmer & Lemeshow, para averiguação da boa adequação do modelo aos dados	56
4.2.8. Análise de possível multicolinearidade entre as variáveis explicativas	57

4.2.9. Análise da tabela das classificações	58
4.2.10. Capacidade de aderência do modelo, R^2 de Nagelkerke do modelo final	60
4.2.11. Linearidade do modelo, uma vez que existe uma variável explicativa contínua	60
4.2.12. Tentativa de ajustamento de regressão logística através de um modelo quadrático	61
4.2.13. Número mínimo de observações para estimação de um modelo Logit	62

Capítulo 5

Aplicação 2 – O caso da escolha dos pneus de Fórmula 1 no Grande Prémio do Mónaco	63
5.1. Introdução – enquadramento do estudo	63
5.1.1. Tipos de pneus usados nas corridas de Fórmula 1	66
5.1.2. Objetivos gerais e específicos do estudo	67
5.1.3. Variáveis utilizadas no estudo	68
5.1.4. Estatística descritiva das variáveis explicativas	71
5.1.5. Metodologia	72
5.1.6. Hipóteses a testar	74
5.2. Resultados obtidos	75
5.2.1. Simulação da probabilidade estimada para cada categoria a diferentes velocidades para carros com e sem apoio aerodinâmico	79
5.2.2. Conclusões dos resultados obtidos	80
5.2.3. Validade do modelo completo	82
5.2.4. Validade do modelo de regressão multinomial que contém apenas as variáveis explicativas significativas	84
5.2.5. Percentagem de previsões corretas do modelo restrito	85
5.2.6. Apresentação das observações em que o modelo está mal ajustado aos dados	86
5.2.7. Simulação em termos de velocidade de ponta	87
5.2.8. Enquadramento das conclusões retiradas	89
5.2.9. Limitações da aplicação	91

Anexos	93
6.1. Aplicação 1	93
6.1.1. Para obter no SPSS o modelo reduzido	93
6.1.2. Para obter a curva ROC do modelo reduzido	95
6.1.3. Para representação da probabilidade de aprovação ao exame de estatística para cada um dos géneros, masculino ou feminino	96
6.1.4. Análise gráfica dos resíduos do modelo restrito	98
6.1.5. Análise de observações influentes que possam afetar a qualidade do modelo	100
6.2. Aplicação 2	104
6.2.1. Modelo restrito	104
6.2.2. Anexos	105
6.3. Tabela Qui-Quadrado unilateral	106
Bibliografia	107

Agradecimentos

Há muitos anos que tenho uma forte paixão e vocação pelo ato de ensinar e investigar nas áreas da estatística e da econometria.

Qualquer atividade que desenvolva nestes domínios é sempre feita com amor incondicional e impulsionada por pessoas que me marcaram ao longo da vida.

Quero em primeiro lugar, referir o meu pai, Honório Rodrigues, a minha tia Manuela e as minhas amigas Madalena e Elizabete, que, pelas condições inerentes à vida humana, já não estão entre nós. Para eles a minha homenagem e o meu agradecimento especial.

Também à minha mãe, Augusta Gomes, ao meu irmão Manuel Gomes, à Margarida Coelho e aos meus sobrinhos, Matilde Coelho e Bernardo Manuel, fonte de energia suplementar, quero dedicar este livro, materialização da minha paixão pelo ensino e investigação. Ainda à Patrícia Soares pela amizade fraterna que me dedica, à Miriam Limeiro, Ana Reis, Paula Ludovina, Filipa Rodrigues, Beatriz Sequeira, Daniela Duarte, Rui Baptista e Diogo Pinto e a alguns professores extraordinários que muito me ensinaram: professor Jorge Andraz e professor Efigénio Rebelo da Faculdade de Economia da Universidade Algarve e aos professores Álvaro Rosa, Margarida Cardoso e Rui Menezes que trabalham no departamento de Métodos Quantitativos do ISCTE. Também não quero deixar de agradecer à Dra. Sónia Leal, diretora do Centro de Saúde de Cascais, e minha médica de família há mais de vinte anos, e ao Dr. Manuel Costa de Matos, diretor do Serviço de Urgências do Hospital de Cascais durante vários anos. Um agradecimento também à Sne, ao Sérgio e ao Dimas da WizArt pela ajuda em termos de utilização de equipamento informático.

Por último, mas não últimos, deixo aqui um abraço terno a todos os estudantes que tive e que muito me ensinaram na consolidação da Arte de ensinar.

Ricardo Alexandre Saraiva Gomes

Cascais, 30 de junho de 2018

Para ser grande, sê inteiro: nada
Teu exagera ou exclui.

Sê todo em cada coisa. Põe quanto és
No mínimo que fazes.

Assim em cada lago a lua toda
Brilha, porque alta vive.

Odes de Ricardo Reis. Fernando Pessoa

Nota introdutória

No decorrer da frequência da parte curricular do programa doutoral em Métodos Quantitativos pela Faculdade de Economia da Universidade de Algarve, apresentei para avaliação de Seminário II de Econometria, um caso aplicado à Regressão com variável dependente com mais de duas categorias de resposta. Tendo em conta a avaliação obtida e o facto de na literatura portuguesa não existir muita diversidade sobre este tema, desejei melhorar e aperfeiçoar o meu trabalho e transformá-lo num livro.

Esta ideia amadureceu quando fui contactado por profissionais que praticam Medicina no Hospital Universitário de Coimbra no sentido de receber apoio na interpretação dos resultados das suas análises em termos de estatística. Com as suas dúvidas e questões apercebi-me que o modo como é apresentado o binómio teoria e prática é muito importante para quem quer aprender e utilizar estas técnicas nos seus estudos e investigações, principalmente em licenciaturas e mestrados nas áreas de Gestão e Medicina e Enfermagem. Sabendo que os modelos de regressão com dados qualitativos como variável dependente são muito utilizados, apercebi-me da falta de um manual prático que servisse de suporte à investigação na família dos modelos generalizados. Estas circunstâncias foram primordiais e incentivadoras para que iniciasse o trabalho que agora apresento aos leitores. De referir que os profissionais atrás citados acabaram por publicar os artigos desejados e manifestaram a satisfação pelo apoio prestado.

Com o advento do desenvolvimento e implementação de *softwares* estatísticos, os investigadores, estudantes e utilizadores destes modelos tão úteis para diminuir a incerteza das suas decisões, por exemplo a nível de banca a concessão ou recusa de um empréstimo é um tema muito peculiar para estes modelos de regressão com variável resposta limitada e qualitativa, viram reduzido o número de horas de trabalho em termos de incrementação no programa estatístico. De facto e após a introdução de uma base de dados e dadas as instruções ao programa por interface ou sintaxe, caso por exemplo do R, a saída dos resultados demora pouco mais de um segundo. A questão é colocada a nível da interpretação dos chamados *outputs*, que podem determinar as melhores decisões se forem feitas de forma correta e assertiva.

Procurei neste livro dar um incremento forte a esta interpretação no aspeto de leitura das diversas equações produzidas por um modelo deste tipo bem como a validação dos diversos pressupostos de avaliação e validação do mesmo, análise de qualidade e ajustamento adequado. Ou seja, tive como preocupação exemplificar de forma exaustiva a especificação, estimação, interpretação de resultados obtidos e diagnóstico desta família de modelos econométricos. Este livro apresenta aspetos teóricos, mas ancorados fundamentalmente na interpretação de resultados obtidos em duas aplicações práticas realizadas com o SPSS, pois foi neste domínio que senti maior dificuldade da parte dos utilizadores, estudantes e investigadores: após a modelação no programa e saída dos respetivos *outputs* não tinham noções estatísticas para o sucesso da respetiva interpretação.

O leitor, neste livro, encontrará a seguir a cada explicação técnica o seu *output* e uma captura de ecrã dos comandos principais do SPSS para modelar o referido tipo de regressão econométrica.

Espero que a consulta deste livro seja útil e esclarecedora para os seus leitores.

Capítulo 1

O modelo de regressão convencional

O modelo de Regressão linear comum, é o modelo em que a variável dependente Y , também chamada de variável endógena, deverá ser métrica e os erros deverão ser independentes e com distribuição Normal $N(0, \sigma^2)$.

Independentemente do número de variáveis explicativas que entram no modelo, quantitativas ou *dummies*, esta exigência é primordial para que a inferência da amostra para a população possa ter qualidade.

Os modelos em que a variável dependente Y é de escolha discreta, quer no caso de existirem duas opções ou mais de duas opções são de extrema importância nos tempos que correm.

Os mesmos possibilitaram a utilização da estatística a situações que requerem suporte de decisão em situações de incerteza; por exemplo na banca, torna-se necessário assegurar com a maior probabilidade possível que um determinado indivíduo está ou não em condições de pagar um empréstimo financeiro concedido pela entidade bancária.

Considere-se o modelo de escolha binária em que a variável dependente toma duas opções; ou seja o atributo terá de ser a escolha de uma de duas opções, através de uma experiência aleatória cujos resultados possam ser modelados por sucesso ou insucesso.

Por exemplo, em termos de uma vontade expressa, posso deslocar-me a um local ou não o fazer, ao candidatar-me a um emprego posso ficar com a vaga disponível ou não o conseguir. Como referi atrás, a banca suscita vários exemplos deste



RICARDO ALEXANDRE SARAIVA GOMES licenciou-se em Gestão de Empresas, especialização Económico-Financeira pela Universidade Moderna de Lisboa, no ano de 2000. Estudou Pós-Graduação em Matemática na qual tomou contato com os primeiros tópicos de Estatística Avançada e Mestrado em Prospecção e Análise de dados no ISCTE-IUL, apresentando a respetiva tese em aplicação de séries cronológicas. Atualmente, frequenta o programa Doutoral em Métodos Quantitativos Aplicados à Economia e Gestão, pela Faculdade de Economia da Universidade de Algarve, especialização em Econometria. Trabalhou na área financeira como Técnico Oficial de contas em empresas como a Grula e Siemens. As funções foram desenvolvidas nas áreas da Contabilidade Geral e Analítica, Gestão Orçamental e Gestão de Projetos. Ensinou no Externato Nacional de Moscavide, foi Monitor de Análise de Dados na Universidade Moderna nos anos de 2007 e 2008 e foi Professor Assistente convidado na Escola Superior de Gestão e Tecnologia do Instituto Politécnico de Bragança. É investigador na área de modelos multivariados de Séries

Cronológicas e Modelos de Regressão com variável dependente qualitativa e limitada, objeto de aplicação na tese de doutoramento que se encontra a preparar. Encontra-se a trabalhar com médicos do Hospital Universitário de Coimbra em termos de apoio estatístico, na publicação de artigos científicos relacionados com as Ciências da Saúde.

Os modelos de regressão econométrica com variável dependente limitada e qualitativa são utilizados com vastos objetivos uma vez que promovem o estudo e conhecimento de novos fenómenos e a tomada de decisão com maior precisão e redução de incerteza, como por exemplo, desde a área das ciências da saúde até à atividade bancária e seguradora, entre outras.

Com uma exposição clara e sucinta da componente teórica, este livro apresenta dois casos práticos, elaborados com o programa SPSS, com a explicação detalhada em termos de interpretação de resultados obtidos, nomeadamente na estimação de parâmetros, validade e diagnóstico de modelos, interpretação dos coeficientes obtidos e ainda qualidade e ajustamento em relação aos dados utilizados.

O primeiro caso prático é feito com a utilização de dados binários e o segundo exemplifica o uso de variável dependente com mais de duas categorias nominais.

A apresentação cuidada das instruções dadas ao programa SPSS e a exposição clara dos resultados obtidos em ambos os casos práticos, contextualizando alguns passos de escrita de um artigo científico na área da estatística, tem como objetivo facilitar a aprendizagem aos leitores que não possuam bases matemáticas muito desenvolvidas.

Este manual destina-se a estudantes do primeiro ciclo do ensino superior de um vasto leque de áreas; no entanto, é ao nível dos mestrados e doutoramentos que a apresentação teórica e prática deste género de modelos é feita de forma mais rigorosa, sendo a sua utilização especialmente indicada. Do mesmo modo, será útil a investigadores e unidades de investigação tendo em conta que estes modelos econométricos são muito utilizados em termos de pesquisa e respetiva apresentação de conclusões e resultados em relatórios e artigos científicos.

Este livro teve o apoio:

