Textos para Discussão
Diretoria de Pesquisas
n.60

Determinantes do Mau Preenchimento da Unidade de Medida na Pesquisa Industrial Anual - Produto

Gustavo Tavares Lameiro da Costa

IBGE
Instituto Brasileiro de Geografia e Estatística

Presidente da República
Jair Messias Bolsonaro

Ministro da Economia
Paulo Roberto Nunes Guedes

Secretário Especial de Fazenda
Bruno Funchal

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

Presidente
Eduardo Luiz G. Rios Neto

Diretora-Executiva
Marise Maria Ferreira

ÓRGÃOS ESPECÍFICOS SINGULARES

Diretoria de Pesquisas
Cimar Azeredo Pereira (em exercício)

Diretoria de Geociências
Claudio Stenner

Diretoria de Informática
Carlos Renato Pereira Cotovio

Centro de Documentação e Disseminação de Informações
Carmen Danielle Lins Mendes Macedo

Escola Nacional de Ciências Estatísticas
Maysa Sacramento de Magalhães

UNIDADE RESPONSÁVEL

Diretoria de Pesquisas

Coordenação de Serviços e Comércio
Alessandro de Orlando Maia Pinheiro

Ministério da Economia
Instituto Brasileiro de Geografia e Estatística - IBGE
Diretoria de Pesquisas
Coordenação de Serviços e Comércio

Textos para Discussão
Diretoria de Pesquisas
número 60

Determinantes do mau preenchimento da Unidade de Medida na Pesquisa Industrial Anual - Produto

Gustavo Tavares Lameiro da Costa

IBGE
Rio de Janeiro
2021

Instituto Brasileiro de Geografia e Estatística - IBGE

Av. Franklin Roosevelt, 166 - Centro - 20021-120 - Rio de Janeiro, RJ - Brasil

ISSN 1518-675X **Textos para discussão - Diretoria de Pesquisas**

Divulga estudos desenvolvidos por técnicos do IBGE e/ou de outras instituições, bem como resultantes de consultorias e traduções consideradas relevantes para o Instituto. A série Textos para discussão está subdividida por unidade organizacional e os textos são de responsabilidade de cada área específica.

ISBN 978-65-87201-55-9

Impressão

Gráfica Digital/Centro de Documentação e Disseminação de Informações - IBGE/CDDI, em 2021.

Capa

Gerência de Editoração - IBGE/CDDI

Costa, Gustavo Tavares Lameiro da
Determinantes do mau preenchimento da unidade de medida na Pesquisa industrial anual – produto / Gustavo Tavares Lameiro da Costa. - Rio de Janeiro : IBGE, 2021.

38 p (Textos para discussão. Diretoria de Pesquisas, ISSN 1518-675X ; n. 60)

ISBN 978-65-87201-55-9

1. Levantamentos industriais - Brasil. 2. Pesquisa Industrial Anual Produto. 3. Estatística industrial - Brasil. 4. Produtividade industrial - Brasil. 5. PRODLIST. I. IBGE. Coordenação de Serviços e Comércio. II. Título. III. Série.

CDU 311.21:338.45(81)

ECO

Sumário

Convenções

-	Dado numérico igual a zero não resultante de arredondamento;
..	Não se aplica dado numérico;
...	Dado numérico não disponível;
x	Dado numérico omitido a fim de evitar a individualização da informação;
0; 0,0; 0,00	Dado numérico igual a zero resultante de arredondamento de um dado numérico originalmente positivo; e
-0; -0,0; -0,00	Dado numérico igual a zero resultante de arredondamento de um dado numérico originalmente negativo.

Resumo

Um dos maiores problemas enfrentados na etapa de coleta dos dados da Pesquisa Industrial Anual - Produto é o preenchimento da unidade de medida que, muitas vezes, não está em conformidade à solicitada pela Pesquisa. Este mau preenchimento pode levar a vários problemas, sendo que o maior deles é a dificuldade na etapa de crítica e imputação dos dados da pesquisa, que tem consumido pelo menos 50% do seu tempo total de apuração. Desta forma, este estudo propõe investigar alguns dos fatores que influenciam no preenchimento do campo das unidades de medida de cada produto (produzido ou vendido) diferente da requisitada na Pesquisa. Os resultados mostram que os fatores mais relevantes são: a alteração na unidade de medida em relação à última edição da lista de produtos da Indústria - Prodlist, o número excessivo de palavras no campo de descrição do produto e a quantidade de códigos NCM associados a um único código de produto da Prodlist. O trabalho mostrou que as alterações no sistema de entrada de dados da Pia-Produto devem ser uma preocupação constante da equipe, incluindo ainda, indicações de mudanças no material de apoio aos informantes, a Prodlist.

Introdução

O problema e os objetivos

A Pesquisa Industrial Anual – Produto (Pia-Produto) é uma pesquisa que tem a coleta iniciada, em geral, no mês de abril e previsão de término em setembro, podendo se estender, extraordinariamente, até dezembro e com divulgação no mês de junho do ano seguinte. A partir do segundo mês de coleta algumas atividades relacionadas à crítica e imputação dos dados começam a ser realizadas.

Cada etapa da pesquisa requer cuidado, sendo que a crítica e imputação de dados necessita maior disponibilidade de tempo e esforço da equipe. A seguir são elencadas algumas destas etapas: codificação dos produtos que não receberam qualquer tipo de classificação; identificação de produtos que porventura foram codificados de forma incorreta por motivo de semelhança em suas características (há casos em que a discriminação entre dois produtos semelhantes é algo difícil); produtos considerados "suspeitos" por terem sofrido grande variação em relação ao seu próprio histórico no tipo de bem (ou serviço) informado ao longo do tempo sem qualquer justificativa; e o considerado foco neste estudo, que é a etapa de verificação do preenchimento das unidades de medida.

O mau preenchimento da unidade de medida desencadeia uma série de dúvidas acerca da informação entregue pelo informante. Para que a definição da unidade de medida seja considerada, alguns protocolos no processo do trabalho são adotados buscando compreender a origem do erro. Há casos em que a unidade de medida respondida é diferente da unidade que a Pia-Produto determina, mas os valores das variáveis[1] entregues são condizentes com valores das respostas de anos anteriores, ou mesmo relativamente parecidos com a média do setor econômico no qual a empresa atua. A decisão, neste caso, está apenas na troca da unidade de medida informada. Há ainda casos onde a unidade de medida é conversível em relação à indicada pelo IBGE, mas há casos em que isso não é possível, como por exemplo, quando uma unidade de medida é definida em massa, mas foi informada em volume[2].

Desta forma, pretende-se ao final deste texto compreender quais fatores influenciam no mau preenchimento das unidades de medida dos produtos da Pia-Produto para que no futuro as equipes possam melhorar os processos de trabalho no momento da crítica da unidade de medida. Este trabalho também poderá orientar a formulação de novos formatos de entrada de dados[3], bem como a criação de críticas automáticas no momento do

.....
[1] Quantidade produzida, quantidade vendida e receita líquida de vendas.
[2] Há produtos cuja variedade de respostas para unidade de medida é grande. Algumas hipóteses podem ser levantadas para que um produto obtenha uma quantidade variada de respostas para unidade de medida. Produtos cujo código de produto abrange vários tipos de produto, como por exemplo, produtos plásticos, têxteis e produtos de metal.
[3] O chamado módulo informante é atualmente disponibilizado em meio eletrônico e preenchido por grande parte dos informantes. Ele já prevê várias críticas automáticas, mas com grande cautela devido à carga de resposta ("response burden").

preenchimento do questionário. *Deve-se salientar que o melhor momento para fazer crítica de dados é na captação da informação*. (Waal, 2011 pg. 214).

Metodologia

Pesquisa Industrial Anual Produto

A Pesquisa Industrial Anual divide-se em duas: uma investiga as variáveis econômicas e outra as ligadas à produção. À primeira dá-se o nome de Pia-Empresa e à segunda Pia-Produto. Esta pesquisa utiliza a informação do estrato certo[4] da Pia-Empresa[5].

Os dados que serão analisados são aqueles que não sofreram nenhuma alteração (crítica) quando da entrada no sistema da Pesquisa. A informação é estudada em sua forma primária, ou seja, é utilizada a base de produtos com dados originais, informada pelas empresas, chamada de Sistema Integrado de Pesquisas Econômicas Anuais - Sipea.

Uma empresa pode ser constituída por várias unidades locais produtivas - ULs, sendo que cada UL pode produzir/vender vários produtos, sendo o conjunto desses produtos chamado de base de produtos. Para a identificação dos principais problemas no preenchimento dos dados da Pia-Produto, serão analisados três anos disponíveis: 2016, 2017 e 2018. A unidade de análise[6] é a informação relativa ao produto. É o resultado da combinação do produto e unidade local onde é produzido ou vendido.

O número das unidades de análise incluídas nesse estudo é menor do que a disponível. Não foram considerados no estudo os serviços industriais, bem como todos os outros produtos que não possuem unidade de medida, como por exemplo, os medicamentos[7]. O número de observações avaliadas, aqui chamadas de linhas, foram as seguintes: 2016 (n'=76.365); 2017 (n'=79.184); e 2018 (n'=76.983). Define-se por linha a combinação de cada UL e cada produto de uma empresa.

Uma outra fonte de informação utilizada é a lista de produtos em que a Pia Produto está baseada. Ela é chamada de Prodlist[8]. É uma lista detalhada de bens e serviços industriais derivada da Classificação Nacional de Atividades Econômicas - CNAE, que possui relação com os critérios adotados internacionalmente (Nomenclatura Comum do Mercosul - NCM e Clasificación Central de Productos - CCP das Nações Unidas).

Optou-se por não utilizar outras bases de dados disponíveis pelo IBGE ou bases externas, uma vez que decidiu-se verificar a capacidade individual da Pia-Produto em

.....
[4] O estrato certo da Pia-Empresa é definido por duas variáveis: Pessoal Ocupado ≥ 30 ou Receita Bruta ≥ rb_t, onde rb_t é o corte de receita bruta para o estrato certo em um ano t. Esse valor é atualizado a cada edição da pesquisa.
[5] Pesquisa Industrial Anual - Empresa: https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=71719 .
[6] Unidade à qual a inferência é dirigida.
[7] O número de observações da Pia-Produto para cada ano: 2016 (n=89.619), 2017 (n=91.249) e 2018 (n=90.622).
[8] https://biblioteca.ibge.gov.br/visualizacao/livros/liv100355.pdf

responder as questões relativas ao mau preenchimento da unidade de medida da pesquisa.

Análise dos Dados – Estatística Descritiva

As análises são feitas por meio das estatísticas descritivas e por modelagem probabilística, levando-se em consideração apenas as variáveis ligadas à Pia-Produto[9]. A abordagem probabilística foi empregada via modelo probabilístico de resposta binária. Optou-se por esta abordagem uma vez que a análise de modelos lineares generalizados com variáveis de respostas binárias reflete a natureza do problema proposto. Desta forma, é considerada a situação na qual a variável resposta possui duas alternativas: (i) resposta dada pelo informante na unidade medida requerida pela Pia-Produto; e (ii) resposta dada pelo informante em unidade de medida diferente à requerida pela Pia-Produto.

Combinando-se a Pia-Produto e a Prodlist é possível tornar a base de dados final capaz de responder as questões relativas ao estudo.

As variáveis investigadas e algumas definições:

- Informante respondeu na unidade de medida exatamente igual à indicada pela Pia-Produto ou; informante não respondeu na unidade de medida indicada pela Pia-Produto – *variável (dummy) resposta* - (RSPUND1 – base de origem = Pia-Produto);
- Tamanho do campo de descrição, traduzido pelo número de palavras, ou o número de caracteres ou o tamanho médio de cada palavra. - (TAMCPDESCR – base de origem = Prodlist 2016);
- Quantidade de códigos NCM associados a um código de produto da Prodlist - (QTNCMS – base de origem = Prodlist 2016);
- Importância da empresa em termos de valor total de vendas em relação à classe econômica – quatro primeiros dígitos do código do produto ($RLV_{UL\ x\ Produto\ k,\ \ Classe\ i}/\sum RLV_{Classe\ i}$) - (PROPCLASSE – base de origem = Pia-Produto);
- Tipos de alterações nos códigos da Prodlist em relação à última edição de 2013, tais como, alteração de: Atividade; Conteúdo; Descrição; Unidade de Medida; NCM; e (Ex)Inclusão – origem = Prodlist 2016;
- Região onde se localiza a Unidade Local produtora;
- Setor e Divisão Econômica[10] – Primeiro e segundo dígitos da CNAE;
- Códigos dos Produtos (cdprods) - são formados por oito dígitos. Os quatro primeiros referem-se à classe CNAE de predominância, o quinto dígito corresponde à CNAE 2.0, onde o algarismo 2 remete às mercadorias e 9 para os serviços industriais; e os três últimos dígitos apenas uma sequência ordenada alfabeticamente dos produtos em cada classe.

.....
[9] Variáveis da Pia-Produto e da Prodlist.
[10] https://concla.ibge.gov.br/classificacoes/por-tema/atividades-economicas/classificacao-nacional-de-atividades-economicas

As Variáveis

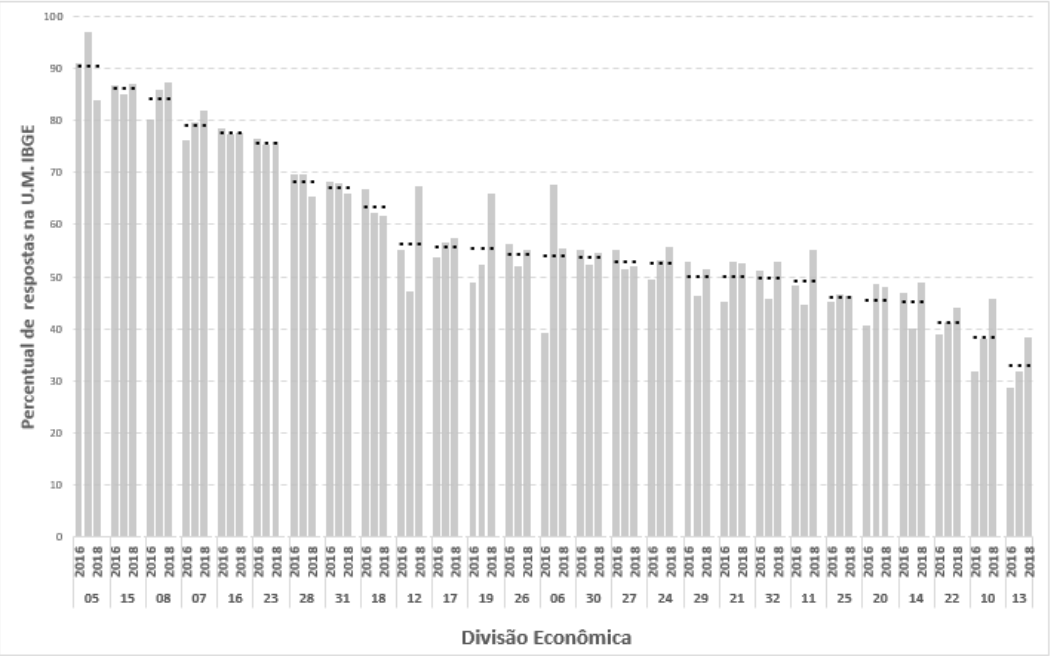
(1) Proporção de respostas na unidade de medida IBGE - RSPUND1

A análise das unidades de medida pode ser feita sob vários aspectos e sob os quais podem ter relação direta com os motivos pelos quais alguns informantes respondem na unidade de medida diferente da exigida pela pesquisa. O gráfico 1 mostra a proporção de informantes que respondeu na unidade de medida solicitada pela pesquisa ao longo dos anos de 2016, 2017 e 2018, segundo a divisão da classificação nacional de atividade econômica. Pode-se verificar que há diferenças importantes quando analisadas estas proporções segundo o setor de atuação da indústria. Os pontos em preto indicam a média aritmética dos três anos para cada divisão. Nota-se que há atividades econômicas cuja proporção de respostas idênticas às requeridas pela pesquisa é muito grande, como a divisão 05 – Extração de Carvão Mineral em contraste à divisão 10 – Produtos Alimentícios cuja média dos três anos ficou um pouco abaixo dos 40%.

É importante notar que, no período de três anos, parte das divisões teve crescimento relativo nas respostas na unidade de medida IBGE (17 das 27 divisões incluídas no estudo). Um possível motivo para a diferença nas proporções médias está em grande parte associado à diversidade de produtos que cada divisão econômica possui, bem como o número de produtores (Tabela 1). A correlação[11] entre a proporção de respostas nas mesmas unidades de medida IBGE e o número de produtos de cada divisão para cada ano foi negativa: $\rho_{2016} = -0{,}3177$, $\rho_{2017} = -0{,}3242$ e $\rho_{2018} = -0{,}4315$. Embora esta relação possa ter sido mais fraca nos primeiros anos, a estatística do teste para correlação ($H_0: \rho = 0 \; x \; H_1: \rho \neq 0$) está ligada ao fato de existirem apenas 27 divisões relacionadas, uma vez que, que alguns setores econômicos não entraram no estudo por se tratarem de serviços. A correlação negativa indica, apenas, que os casos em que as divisões econômicas obtiveram uma proporção de respostas na unidade de medida IBGE consideradas alta foram as mesmas cuja quantidade de produtores foi menor[12]. Fato interessante a ser destacado é que a correlação se acentua ao longo do tempo. Esta variável será fundamental para que se compreenda o problema uma vez que for incluída no modelo probabilístico proposto.

.....
[11] Para 2016 e 2017 a estatística do teste foi significativa com p-valor abaixo de 0,1 e em 2018 foi significativa com p-valor de 0,02. A correlação de Pearson foi calculada.
[12] A correlação indica apenas a associação das variáveis e de forma alguma uma relação de causalidade que sempre deve ser testada probabilisticamente.

Gráfico 1. Percentual de respostas na unidade de medida requisitada pela Pesquisa Industrial Anual Produto, segundo as divisões econômicas, anos e a indicação da média dos três anos, 2016, 2017 e 2018



Fonte: Pesquisa Industrial Anual Produto, 2016, 2017 e 2018.

(2) Tamanho do campo de descrição do produto na Prodlist – TAMCPDESCR

Outra questão sempre comentada e destacada pela equipe da Pia-Produto é o fato de alguns informantes não encontrarem a descrição dos produtos exatamente com as características descritas na Prodlist. Apesar de não haver no questionário nenhum campo que indique se um produto foi encontrado com características exatas na Prodlist é possível que por meio de uma variável *proxy* tal informação seja captada. O tamanho do campo de descrição do produto indicará se de fato ele influencia na decisão de se escolher a unidade de medida idêntica à exigida pela pesquisa ou não.

A tabela 1 mostra que o tamanho médio de palavras, caracteres e caracteres por palavra no campo de descrição varia consideravelmente entre as divisões de classificação econômica. A tabela compara a informação constante nos campos da descrição da Prodlist com os que foram efetivamente encontrados na amostra. As diferenças entre os dois valores estão no fato de alguns produtos terem sido produzidos ou vendidos em maior quantidade em relação a outros dentro da mesma divisão.

Tabela 1. Valor médio do total de palavras, caracteres e caracteres por palavras na Pia-Produto

Divisão	Nº Produtos	Total de Palavras (média)		Total de Letras (média)		Total de Letras por Palavra	
		Prodlist	Pprod (amostra)	Prodlist	Pprod (amostra)	Prodlist	Pprod (amostra)
05	2	14,5	11,7	95,5	86,0	6,9	7,6
06	4	6,8	5,6	37,8	29,5	5,6	5,3
07	19	12,6	14,1	72,5	89,1	5,7	6,2
08	47	8,6	8,8	51,2	51,9	6,3	6,2
10	303	8,6	9,1	49,4	53,3	5,8	6,0
11	31	7,2	6,9	42,9	42,8	6,3	7,3
12	7	9,4	12,6	56,9	80,4	6,3	6,7
13	131	11,5	11,7	67,9	70,4	5,9	6,1
14	85	9,8	9,1	57,8	53,4	5,8	5,7
15	62	12,9	15,5	77,0	90,8	6,0	5,9
16	42	10,1	9,2	61,1	57,3	6,0	6,2
17	85	9,8	9,5	62,0	58,6	6,6	6,2
18	41	10,7	12,9	71,4	85,8	6,8	6,8
19	46	8,1	12,5	48,3	70,3	6,6	6,2
20	477	7,4	8,2	46,8	51,1	6,8	6,6
21	34	7,5	10,1	52,5	70,6	7,9	7,1
22	111	10,1	11,3	62,3	68,5	6,3	6,0
23	110	11,4	11,5	67,2	72,1	6,0	6,2
24	112	9,8	10,7	53,2	56,2	5,9	5,5
25	182	10,8	12,7	64,5	74,3	6,0	6,0
26	156	9,8	10,7	63,1	68,4	6,9	6,8
27	134	9,9	11,3	59,6	67,9	6,2	6,2
28	355	9,6	10,1	59,2	62,6	6,5	6,4
29	79	13,1	11,9	79,2	72,7	6,4	6,3
30	60	10,7	12,5	64,5	72,1	6,5	5,8
31	62	8,5	9,6	46,7	54,3	5,4	5,6
32	143	9,9	10,2	62,7	65,1	6,4	6,4

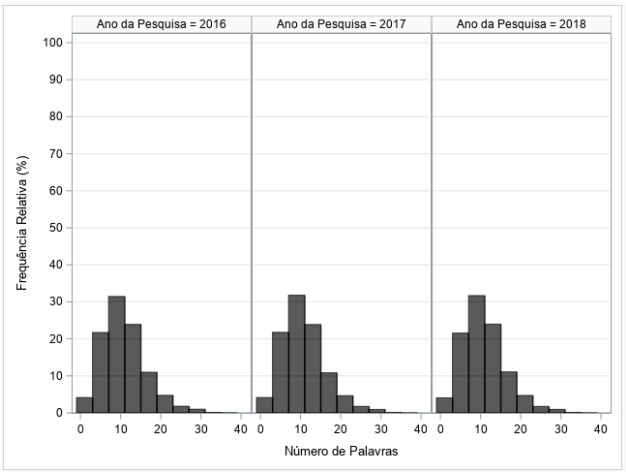
Fonte: Pesquisa Industrial Anual Produto, 2016, 2017 e 2018.

Pode-se notar que há divisões em que são exigidas poucas palavras para a descrição do produto como a divisão 06 – Extração de Petróleo e Gás Natural em contraposição à divisão 29 – Veículos Automotores que precisam, em média, treze palavras para poder descrever seus produtos. Como era de se esperar, os valores para a Pesquisa são parecidos nos três anos, destacados nos histogramas do gráfico 2. Por isso, optou-se apresentar na tabela 1 apenas a média aritmética dos três anos. Isto acontece, pois, grande parte das unidades locais produtivas fabrica os mesmos produtos ao longo do tempo, o que pode ser verificado analisando-se a distribuição empírica do tamanho do campo de descrição.

Contrastando o tamanho do campo de descrição com a característica do mau preenchimento do campo da unidade de medida nota-se que, em geral, as diferenças parecem não existir entre os dois grupos de unidade de medida (gráfico 3). No entanto, para algumas divisões econômicas esta diferença é aparente[13] (gráfico 4) e para outras muito sutis (gráfico 5). A variável que se mostrou melhor para o TAMCPDESCR foi o número de palavras do campo de descrição do produto.

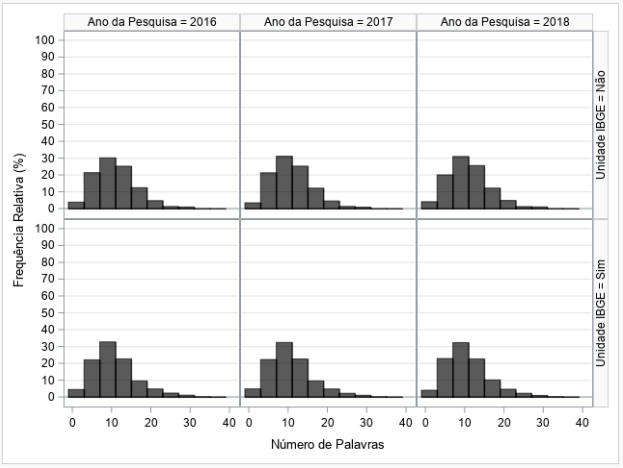
.....
[13] No entanto, um teste estatístico deve ser realizado para a constatação probabilística.

Gráfico 2. Distribuição do nº de palavras de todos os produtos para cada um dos três anos da pesquisa



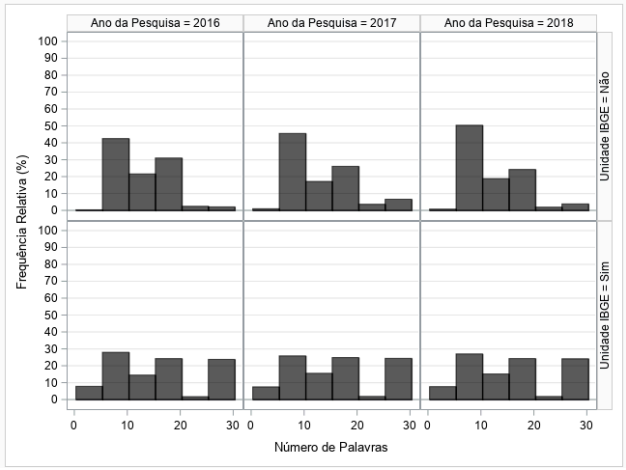
Fonte: Pesquisa Industrial Anual Produto, 2016, 2017 e 2018.

Gráfico 3. Distribuição do nº de palavras, segundo o tipo de unidade de medida para cada um dos três anos da pesquisa



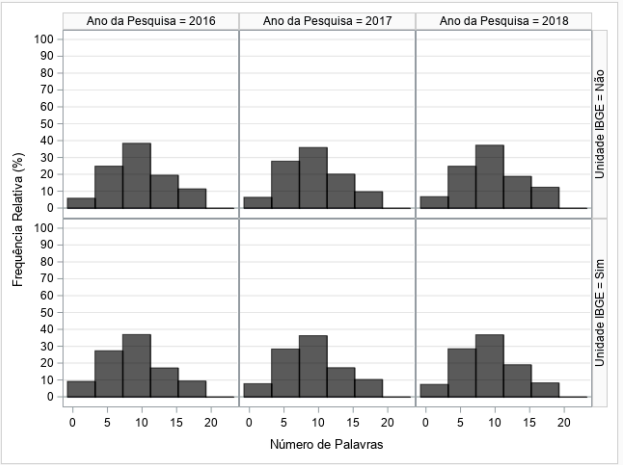
Fonte: Pesquisa Industrial Anual Produto, 2016, 2017 e 2018.

Gráfico 4. Distribuição do nº de palavras da CNAE 15, segundo o tipo de unidade de medida e anos da pesquisa



Fonte: Pesquisa Industrial Anual Produto, 2016, 2017 e 2018.

Gráfico 5. Distribuição do nº de palavras da CNAE 17, segundo o tipo de unidade de medida e anos da pesquisa



Fonte: Pesquisa Industrial Anual Produto, 2016, 2017 e 2018.

Com o exposto, a hipótese de inclusão de investigação desta variável em um modelo probabilístico pode ser fundamentada no fato de quanto maior for o campo de descrição do produto mais assertivo será o informante em relação a encontrar o seu produto na lista de produtos – Prodlist. Após testar as três opções apresentadas na Tabela 1, optou-se por utilizar o número de palavras no campo de descrição.

(3) Quantidade de NCMs associada ao Código de Produto – QTNCMS

Os códigos dos produtos (cdprods) da Pia-Produto – Prodlist possuem relação com os códigos atribuídos aos produtos da Nomenclatura Comum do Mercosul – NCM. A primeira edição da Prodlist, em 1998, baseou-se na lista de aproximadamente 10.000 códigos NCM para que naquele ano fosse criada a lista com 4.672 produtos. A edição da Prodlist 2016 contém 3.400 produtos para 13.051 códigos NCM. Pelo fato da quantidade de códigos NCM ser maior que a de cdprods alguns destes englobam muitos códigos NCM. Várias empresas baseiam-se nos códigos NCM para identificação de seus produtos,

pois o enfoque delas pode estar na produção para vendas ao exterior e por isso a correspondência biunívoca com os códigos da Prodlist é o ideal. Pelo fato dos cdprods serem mais agregados que os da NCM, o informante pode localizar com menos facilidade o código do produto.

A tabela 2 mostra de forma consistente um valor maior para todos os três anos na quantidade média de códigos NCM para os casos nos quais onde a unidade de medida foi respondida diferentemente da unidade IBGE. Os intervalos de confiança para a quantidade de NCMs média incluídos na tabela foram calculados ao nível de confiança de 95%. Em nenhum dos anos há interseção entre os *ICs* indicando que a diferença entre eles é estatisticamente significativa.

Tabela 2. Quantidade de NCMs associadas aos códigos de produtos da Prodlist, segundo classificação da unidade de medida - nº médio e Intervalos de Confiança (ICs), 2016 a 2018

Resposta na U.M. IBGE ?	2016			2017			2018		
	média	LI	LS	média	LI	LS	média	LI	LS
Igual	4,07	4,01	4,13	4,09	4,03	4,16	4,26	4,2	4,32
Diferente	4,76	4,68	4,84	4,8	4,73	4,88	4,62	4,55	4,7

Fonte: Pesquisa Industrial Anual Produto, 2016, 2017 e 2018.

O fato da resposta indicada por um informante ser por vezes em uma unidade de medida diferente da esperada pode estar associado ao número de códigos NCM para um único código Prodlist. Desta forma, esta variedade de códigos NCM pode causar um "confundimento" no momento do preenchimento da pesquisa por parte do respondente, não sabendo ao certo a unidade de medida a ser preenchida. Por este motivo esta variável fará parte do modelo estocástico proposto.

(4) Participação da Receita Líquida de Vendas em relação ao total da classe econômica - PROPVV

Outra variável relevante para a análise do preenchimento incorreto da unidade de medida por parte do informante é a participação no total da receita líquida de vendas de cada produto na classe da classificação econômica (4 dígitos do código de produto). A hipótese diz respeito à ideia de que quanto maior a participação, mais representativos serão os produtos em relação à classe onde se encontram e, por consequência, serão produzidas por unidades locais cujas empresas são importantes para o setor que representam. A utilização desta variável neste formato foi escolhida, em vez da Receita Líquida de Vendas em formato absoluto, para que a comparação não fosse prejudicada entre as observações em um mesmo ano ou entre anos.

A tabela 3 indica diferenças importantes entre as proporções médias da Receita Líquida de Vendas em cada classe econômica segundo a classificação da unidade de medida dada pelo informante em relação a sugerida pelo IBGE. Para todos os anos a participação da RLV foi maior nos casos da unidade de medida IBGE – Tabela 3 – (0,40/0,27=1,46 (2016); 1,44 (2017); e 1,76 (2018)). O teste *t* para amostras independentes foi realizado e a diferença foi estatisticamente significativa para os três

anos ao nível de confiança de 95% comparando-se as médias de PROPVV nos casos das unidades de medida serem ou não entregues na exigida pela pesquisa.

Tabela 3. Valor médio e coeficiente de variação da participação do valor de vendas do produto em sua classe econômica, segundo a classificação da unidade de medida, Pia-Produto, 2016 a 2018

Resposta na u.m. IBGE?	2016		2017		2018	
	média	cv	média	cv	média	cv
Igual	0,40	492,4	0,38	523,8	0,42	515,7
Diferente	0,27	547,9	0,27	576,8	0,24	496,6

Fonte: Pia-Produto, 2016, 2017 e 2018.

(5) Alteração nos códigos de produto: edição da Prodlist 2016 comparada à 2013

Diferentemente das primeiras edições da Prodlist para a Pesquisa Industrial Anual – Produto que teve início em 1998, uma nova edição[14] é publicada para que sirva de base para três anos da pesquisa. A edição utilizada neste trabalho foi a de 2016. Ela foi utilizada na coleta da Pia-Produto para os anos de 2016, 2017 e 2018.

A cada nova edição são elencados os produtos que sofreram algum tipo de alteração nos seus códigos. Isto acontece por diversos motivos: novos produtos surgem em diversos setores, enquanto outros desaparecem, como a televisão de tubo, substituída pelas de tela plana. Outros deixaram simplesmente de ser fabricados, como a secretária eletrônica. Há ainda os casos em que houve a necessidade de agregações e desagregações solicitadas pelos próprios produtores via associações.

A rotina de manutenção e revisão segue alguns critérios de atualização. *Inclusão* (I) – novos produtos/serviços que não constaram em edições anteriores; *Exclusão* (E) – retirada de produtos/serviços não mais produzidos ou sem qualquer justificativa econômica; *Agregação* (AG) – casos em que os produtos têm descrições detalhadas, que por vezes dificultam sua identificação; produtos que não possuem justificativa econômica de existirem; ou produtos com descrições duplicadas; *Alterações da descrição* (AD) – alteração na ordem das palavras bem como consertos na grafia ou atualização de algum termo que se tornou obsoleto; *Alterações de conteúdo* (AC) – sofrem mudança em relação à abrangência do conteúdo. Nesses casos, são alterados os códigos de origem; e *Alterações de informações vinculadas* – abrangem as modificações da unidade de medida (AM) adotada como padrão, bem como outras correspondências com outros sistemas de classificação com os quais a Prodlist da Indústria está embasada, como por exemplo, a alteração na NCM (AN) e alteração de atividade ou código da Prodlist (AA).

As tabelas A.3 e A.4, incluídas no anexo, mostram as proporções de casos para cada ano, segundo seus tipos de alteração. Salienta-se que pelo fato das amostras entre os anos serem parecidas os valores não sofrem grandes alterações. A tabela A.3 indica

[14] Edições da Pia-Produto - 1998-2007; 2010; 2013; 2016; e 2019.

alterações individuais: em destaque estão as unidades de medida e descrição. A tabela A.4 indica que a combinação de dois tipos ou mais de alteração, no entanto, não foram testados no modelo estas combinações.

Modelagem Probabilística

Geralmente, em textos técnicos, a modelagem probabilística só é realizada após a estatística descritiva. O ideal é que esta análise seja feita em contraste com a variável resposta do modelo probabilístico, que neste caso é a variável relativa às respostas das unidades de medida dos produtos solicitadas pela pesquisa (RSPUND1). No capítulo anterior foram apresentadas as variáveis que poderão compor o modelo final, no entanto, a inclusão delas no modelo final só fará sentido caso tenham alguma associação à variável resposta.

Além desse fato, os possíveis efeitos entre as variáveis devem ser levados em consideração antes de integrarem o modelo final. Ou seja, variáveis que possuem forte correlação[15] estatística não precisam e não devem compor o mesmo modelo probabilístico. Esta condição recebe o nome de multicolinearidade[16]. Este problema é constatado quando há pelo menos duas covariáveis (ou fatores) do modelo probabilístico correlacionadas. Por este motivo as covariáveis correlacionadas foram cuidadosamente testadas de diversas formas para que se chegasse ao formato final do modelo. A forma mais fácil de se corrigir problemas de multicolinearidade é retirando as variáveis correlacionadas a critério do pesquisador e neste trabalho foi seguida a ordem de importância das variáveis.

A inclusão indiscriminada de variáveis (inclusive as indicadoras) em um modelo probabilístico pode causar outro problema: o efeito de “confundimento” nos parâmetros estimados, com sinais diferentes dos esperados (Achen C., 1985). Muito embora, o modelo final deste estudo possui um número reduzido de variáveis frente ao número de observações disponíveis, este problema, ainda assim, poderia ter acontecido caso não tivessem sido tomados os devidos cuidados.

A tabela A.2 incluída no anexo apresenta a matriz de correlação com todas as variáveis que poderão ser incluídas no modelo final. É possível identificar que algumas relações entre variáveis não são consideradas significativas do ponto de vista estatístico e isso é considerado algo positivo para modelagem. Um exemplo é o número de palavras (TAMCPDESCR) e a participação do valor da receita líquida de vendas de um produto em relação ao total de sua classe (PROPVV). Isto acontece também quando TAMCPDESCR é comparado ao valor de vendas absoluto para os três anos. Nota-se, também, que TAMCPDESCR e as variáveis que envolvem o valor de vendas não são correlacionadas sob qualquer formato.

.....

[15] A correlação de Pearson foi utilizada: $\rho = \sum_k (x_k - \bar{x})(y_k - \bar{y}) / \sqrt{\sum_k (x_k - \bar{x})^2} \sqrt{\sum_k (y_k - \bar{y})^2}$

[16] Quando isso ocorre o modelo não se torna mais “inversível” por não ser mais de posto completo.

Por outro lado, como apontado anteriormente, o problema está no fato de existirem variáveis elencadas para integrarem o modelo probabilístico, cuja correlação é forte. Estes casos, restringiram-se às variáveis sem qualquer tipo de transformação (matemática ou não) e suas respectivas transformações (logarítmica). Dessa forma, optou-se por utilizar as variáveis não transformadas uma vez que elas são de fácil interpretação e assimilação. A variável PROPVV tem correlação moderada com o valor de vendas em milhões de reais. Pelo fato da participação ser considerada uma variável importante por ser calculada relativamente à cada divisão econômica, os modelos com ou sem esta variável foram avaliados e optou-se por incorporá-la definitivamente ao modelo final. Foram testados modelos com ambas as variáveis. Verificou-se que a inclusão do valor absoluto de vendas ao último modelo considerado foi significativa do ponto de vista probabilístico[17].

O Modelo

Um modelo probabilístico adequado para que se compreenda o mau preenchimento da unidade de medida dos produtos na Pia-Produto é o que considera como variável resposta apenas duas categorias. O modelo de regressão logística é um dos que tem essa característica, uma vez que a interpretação de seus parâmetros pode ser feita de forma direta e fácil. Ele é oriundo de uma categoria de modelos chamada modelos lineares generalizados para um componente aleatório binário com a função de ligação, ***logit***, diferente do modelo de mínimos quadrados (MMQ) que é a distribuição normal. A ideia central é modelar a probabilidade de sucesso, $P(U.M.Informada = U.M.IBGE|\boldsymbol{X})$ onde $\boldsymbol{X} = (x_1, x_1, \dots x_v)$ é um vetor de v variáveis independentes, que neste trabalho foram tratadas como covariáveis. A probabilidade condicional para se obter a informação na unidade de medida de acordo com a exigida pelo IBGE segundo algumas covariáveis pode ser simplificada pelo termo $\pi(x)$.

O modelo logístico é definido por:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (1)$$

E aplicando-se a transformação logarítimica:

$$g(x) = ln\left[\frac{\pi(x)}{1 - \pi(x)}\right] \quad (2)$$

Obtém-se a função logito que é dada por:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_v x_v \quad (3)$$

[18]

Deve-se compreender que a relação da proporção $\pi(x)$ e uma covariável qualquer é não linear, diferentemente dos MMQ. Por causa do formato em S da distribuição

[17] Aqui foi utilizado o critério da *Deviance.*

[18] O modelo correto a ser descrito deve conter o termo exclusivo para as variáveis indicadoras D_{jl} ($\sum_{l=1}^{k_j-1} \beta_{jl} D_{jl}$). No entanto, foi suprimida esta parte por motivos de simplificação do modelo.

logística, uma mudança em x_k pode ter menos impacto quando $\pi(x)$ está perto de zero ou da unidade do que quando $\pi(x)$ está concentrado nas partes centrais da distribuição x . No entanto, isto não é um problema, uma vez que o modelo de regressão logística tem a forma linear para o *logito* da probabilidade de sucesso, como destacado na equação (2).

O efeito dos parâmetros estimados $(\widehat{\beta_1}, \widehat{\beta_2}, \dots, \widehat{\beta_v})$ determina a taxa de crescimento ou decrescimento da curva em formato de S para a probabilidade de o informante entregar a informação na unidade de medida IBGE, ou seja, a probabilidade de sucesso.

O modelo logito possui algumas importantes características (ver Gujarati, D. 2000):

1. Conforme P varia de 0 a 1 (isto é, conforme Z varia de $-\infty$ a$+\infty$), o logito $g(x)$ vai de $-\infty$ a$+\infty$. Ou seja, embora as probabilidades se situem entre 0 e 1, os logitos não se restringem a esses limites e por isso necessitam de serem transformados para serem compreendidos;
2. Embora o logito seja linear em x, as probabilidades propriamente ditas não são por conta da distribuição de probabilidades atribuída ao modelo logístico;
3. Interpretam-se os parâmetros estimados, exceto o intercepto, como a inclinação que mede a variação do logito para uma mudança unitária na covariável. Ela diz, por exemplo, como a chance, em logaritmo, em favor da resposta ser na unidade de medida IBGE varia a cada milhão de reais a mais no valor de receita líquida de vendas de um produto;

Ressalta-se que o enfoque deste trabalho está na intepretação dos modelos e como eles podem auxiliar na elaboração de protocolos futuros para melhorias no sistema de coleta da informação da Pia-Produto. Aqui não serão apresentadas a teoria referente à regressão logística em termos dos cálculos de estimação de variância, assim como dos intervalos de confiança, que para o modelo logístico é um pouco diferente do método dos mínimos quadrados. Para mais informações ver Hosmer e Lemeshow (2013).

Análise dos Dados – modelo probabilístico

Variável Resposta

O modelo retrata o objetivo do estudo propondo utilizar a variável dicotômica: unidade de medida selecionada pelo informante igual à indicada pela Pia-Produto (indicador=1) e unidade de medida diferente à indicada pela Pia-Produto (indicador=0). Utilizando os fatores corretos é possível entender quais são os determinantes do mau preenchimento da unidade de medida pelos informantes da Pia – Produto.

Covariáveis (Fatores)

Alguns modelos foram testados incluindo as variáveis descritivas apresentadas. Parte deles apresentou variáveis que não foram importantes ou geraram algum tipo de “confundimento” na interpretação do modelo. As variáveis que tiveram relação importante com a variável resposta foram:

Tabela 4. Variáveis incluídas no modelo logístico

Variável	Tipo da Variável	Valores
Alteração da Unidade de Medida	Indicadora	1- Alterada; 0 – Não alterada
Alteração de Conteúdo	Indicadora	1- Alterada; 0 – Não alterada
Alteração de Descrição	Indicadora	1- Alterada; 0 – Não alterada
Alteração de NCM	Indicadora	1- Alterada; 0 – Não alterada
Nº de NCMs associadas a um código Prodlist	Numérica – Discreta	
Valor de Vendas em R$ milhões	Numérica – Contínua	
Nº de palavras no campo da descrição da Prodlist	Numérica – Discreta	
Participação das vendas na classe de produto	Numérica – Proporção	
Variáveis de controle: Região e Seção Econômica	Indicadoras	Região (ref. SE) e Seção (ref. 1)

Fonte: Pia Produto -2016 a 2018 e Prodlist 2016

Algumas variáveis que não se alteram ao longo do tempo, tais como a atividade econômica e o local da unidade de coleta servem como variáveis controle para a modelagem probabilística e devem ser levadas em consideração. O local de coleta é incluído no modelo por meio do código da região de coleta e seus coeficientes estimados são comparados à região de referência. Neste modelo a região Sudeste foi considerada como a de referência.

A inclusão do setor econômico do modelo foi pelo primeiro dígito do código de produto. Esta decisão é baseada no fato de que os coeficientes estimados para cada divisão econômica não serão utilizados no trabalho, no entanto, essa informação deve ser considerada de alguma forma na estrutura do modelo.

Variáveis que têm relação com as características intrínsecas aos produtos, tais como, o tamanho do campo de descrição de cada produto e os diversos tipos de alterações nos códigos dos produtos, que só sofrerão quaisquer mudanças quando a lista de produtos for revisada, também são covariáveis que não variam no tempo. Ou seja, enquanto não houver reedição da lista de produtos – Prodlist o seu conteúdo é o mesmo. No entanto, diferentemente das variáveis controle, estas têm correspondência em relação à edição da publicação anterior, a Prodlist 2013.

A utilização das informações da Prodlist (uma edição em comparação à anterior) agrega informação relevante ao modelo, pois cada observação estabelecerá uma relação com as mudanças referentes à publicação passada, muito embora, com resultados parecidos para os diferentes anos por conta da interseção das observações ser importante[19].

.....

[19] A sobreposição para os três anos para cada nível da informação foi: Empresa 71,2%; Unidade Local 70%; e Produto 47,8%.

Resultados

Um importante fator levado em consideração na análise dos dados foi o tempo, uma possível variável explicativa na discussão deste estudo. Foi incluído no anexo uma explicação para a consideração do tempo na análise dos dados. Como havia disponibilidade de utilização da informação para os anos de 2016 a 2018, optou-se inicialmente por utilizá-los comparando os resultados por meio das estatísticas descritivas e, por conseguinte os modelos probabilísticos separados para cada ano. Inicialmente foi vislumbrada a possiblidade de se utilizar o tempo como covariável no modelo, no formato de modelo de dados em painel. Todavia, nos casos dos modelos de resposta binária a interpretação dos resultados, assim como o processo de estimação dos parâmetros[20] tornam-se obstáculos por conta do tamanho da base de dados e da performance do software[21].

A proposta feita neste texto é de comparar os resultados estimados para cada um dos anos. Esta comparação pode ser feita uma vez que grande parte das observações é a mesma. Apesar de não ser um painel fixo, as empresas, unidades locais e produtos comparados são oriundos do estrato certo da Pesquisa Industrial Anual Empresa. Soma-se a este fato a característica de o setor industrial ser pouco variável a curto prazo, muito por conta de que pelo menos um fator de produção ser fixo (ver Kupfer, D. e Hasenclever, L. 2013, pág. 5) e, assim, os resultados tornam-se confiavelmente comparáveis[22] uma vez que o efeito da variação das unidades observacionais é considerado importante como fator secundário.

Cabe ressaltar que algumas variáveis foram testadas em formatos distintos. O modelo descrito na Tabela 5 utilizou o formato original das variáveis, e alternativamente foram incluídas variáveis com transformações matemáticas, ou com categorias ordenáveis em um segundo modelo (Tabela 7). Estas tentativas foram importantes para que fossem verificadas diferenças em diferentes níveis das covariáveis quando da interpretação dos coeficientes estimados à luz da variável resposta. Nota-se na Tabela 5 que alguns parâmetros estimados foram significativos apenas em alguns anos e o importante é compreender a evolução dos parâmetros estimados por meio da análise da variação dos sinais. Pode-se entender, por exemplo, se houve melhora ou não em relação à adequação dos informantes às unidades de medida da Pia Produto segundo a importância relativa da receita líquida de vendas (PROPV) na classe cujo produto está inserido.

As três variáveis indicadoras que carregam a informação de alteração do código dos produtos tiveram comportamentos distintos como se pode notar na Tabela 5. Deve-se lembrar que estas variáveis comparam a situação da edição da Prodlist 2016 com a de

.....

[20] O modelo misto de efeito aleatório adequado neste caso utilizaria a Proc NLmixed do software SAS. Pelo fato de usar iterações numéricas o modelo não executou.

[21] Modelos de dados em painel, de efeito fixo ou aleatório, são muito difíceis de executados com grandes bases de dados. Os modelos de efeito aleatório requerem suposições muito fortes sobre a heterogeneidade do resíduo do modelo e os de efeitos fixos precisam de alterações no cálculo do estimador de máxima verossimilhança que se nada for feito permanece inconsistente (Greene, 2002).

2013. No caso da alteração na unidade de medida causou uma piora na resposta na unidade de medida indicada pelo IBGE, muito embora, o aumento do valor dos parâmetros estimados indica que ao longo do tempo essa alteração na unidade de medida entre melhorou apesar de ainda apresentar um efeito negativo. A primeira mostra que com o aumento do parâmetro estimado os informantes que tinham em sua lista produtos que sofreram alguma alteração na unidade de medida em relação à 2013 passaram entregar a informação na unidade de medida correta. Em 2016, na tabela 8, verifica-se que, em média, a proporção foi de 27% ($= e^{-1,31}$) e em 2018 essa passou para 37%.

Tabela 5. Resultados do modelo de regressão logística com as variáveis importantes, para cada ano do estudo

Variável	Nível da Variável	Par.Est. 2016	Par.Est. 2017	Par.Est. 2018
Intercepto		1,019*	1,113*	1,08*
Alteração U.M.	1 = alterada	-1,31*	-1,252*	-1,001*
Alteração Descrição	1 = alterada	0,263*	0,253*	0,244*
Alteração NCM	1 = alterada	-0,833*	-0,485*	-0,803*
Nº palavras	variável discreta	-0,02*	-0,021*	-0,018*
Participação Classe	proporção	0,077*	0,063*	0,083*
Participação Classe^2	termo quadrático	-0,001*	-0,001*	-0,001*
Nº de códigos NCM	variável discreta	-0,008*	-0,008*	-0,001
Valor de Vendas (R$ milhões)	variável contínua	-0,00009*	0,000005	0,0004*
Região	Norte	-0,134*	-0,227*	-0,108*
(ref= Região Sudeste)	Nordeste	0,229*	0,177*	0,264*
	Sul	-0,095*	0,027	-0,114*
	Centro Oeste	0,011	0,085*	-0,005
Seção	Zero	0,799*	1,137*	1,011*
(ref=seção 1)	Dois	-0,174*	-0,21*	-0,322*
	Três	0,038	-0,172*	-0,166*

Fonte: Pesquisa Industrial Anual – Produto, 2016, 2017 e 2018.

Nota: * coeficientes estimados significativos ao nível de significância de 5%.

O número de palavras no campo da descrição do código do produto Prodlist afetou negativamente a probabilidade dos informantes responderem na unidade de medida IBGE. Quanto maior a participação da Receita Líquida de vendas em cada classe de produto melhor é a entrega da informação na unidade de medida exigida pela Pesquisa. Ou seja, os informantes cuja importância relativa for grande em sua classe[23], em termos de RLV do produto melhor será a resposta na unidade de medida IBGE. No entanto, essa importância tem um limite indicado pelo termo quadrático negativo da mesma variável. O fato dos coeficientes estimados para PROPV serem muito parecidos para os três anos mostra que não houve mudanças importantes para esses respondentes ao longo do tempo. Ou seja, eles continuam entregando a informação de forma correta.

A quantidade de códigos NCM referente a cada código de produto Prodlist está negativamente associada a entrega da resposta na unidade de medida IBGE. Produtos com vários códigos NCM têm respostas piores que produtos cuja quantidade de códigos

[23] Os quatro primeiros códigos do cdprod.

NCM é menor. Analisando-se os coeficientes no tempo parece que em 2018 esses produtos apresentaram uma pequena melhora em suas respostas.

Tabela 6. Razão de Chances referentes ao modelo logístico com variáveis originais

Variável	Nível da Variável	Estimativa 2016	Estimativa 2017	Estimativa 2018
Alteração U.M.	1 = alterada	0,27 (0,26;0,28)	0,29 (0,28;0,3)	0,371 (0,359;0,383)
Alteração Descrição	1 = alterada	1,3 (1,22;1,39)	1,29 (1,21;1,37)	1,283 (1,204;1,367)
Alteração NCM	1 = alterada	0,44 (0,36;0,53)	0,62 (0,52;0,74)	0,449 (0,375;0,537)
Nº palavras	variável discreta	0,98 (0,98;0,98)	0,98 (0,98;0,98)	0,982 (0,979;0,985)
Participação Classe	proporção	1,08 (1,06;1,1)	1,07 (1,05;1,08)	1,087 (1,067;1,107)
Nº de códigos NCM	variável discreta	0,99 (0,99;1)	0,99 (0,99;0,99)	0,99 (0,99;1,001)
Região	Norte	0,89 (0,8;0,98)	0,85 (0,77;0,94)	0,938 (0,853;1,032)
(ref= Região Sudeste)	Nordeste	1,27 (1,2;1,34)	1,27 (1,2;1,34)	1,359 (1,287;1,435)
	Sul	0,92 (0,89;0,95)	1,09 (1,06;1,13)	0,932 (0,901;0,964)
	Centro Oeste	1,02 (0,94;1,11)	1,16 (1,07;1,25)	1,04 (0,962;1,125)
Seção	Zero	4,32 (3,8;4,9)	6,63 (5,75;7,65)	4,643 (4,018;5,364)
(ref=seção 1)	Dois	1,63 (1,58;1,69)	1,72 (1,67;1,78)	1,214 (1,176;1,253)
	Três	2,02 (1,89;2,15)	1,79 (1,68;1,91)	1,431 (1,346;1,521)

Fonte: Pesquisa Industrial Anual – Produto, 2016, 2017 e 2018.

Nota: * coeficientes estimados significativos ao nível de significância de 5%.

A melhor forma de se analisar os resultados de uma regressão logística é por meio das razões de chances. Os resultados estão apresentados na tabela 6. As razões de chances comparam a probabilidade de sucesso à probabilidade de fracasso.

A razão de chances para o ano de 2016 no exemplo 1 da tabela 7 é calculada da seguinte forma:

$$P(u.m = IBGE|X) = 1{,}019 - AUM \cdot 1{,}31 + ADES \cdot 0{,}263 - ANCM \cdot 0{,}833 - \text{TAMCPDESCR} \cdot 0{,}02 + PROPV \cdot 0{,}077 - PROPV^2 \cdot 0{,}077 - QTDENCMS \cdot 0{,}008 - VVMI \cdot 0{,}00009 - REGN \cdot 0{,}134 + REGNE \cdot 0{,}229 - REGS \cdot 0{,}095 + REGCO \cdot 0{,}011 + SECAO0 \cdot 0{,}799 - SECAO2 \cdot 0{,}174 + SECAO3 \cdot 0{,}038$$

$$\hat{g}(AUM = 0, ADES = 0, ANCM = 0, TAMCPDESCR = 5, PROPV = 5, QTDENCMS = 1, VVMI = 15, REGIAO = SE, SECAO = 1)$$
$$= 1{,}019 - 0 \cdot 1{,}31 + 0 \cdot 0{,}263 - 0 \cdot 0{,}833 - 5 \cdot 0{,}02 + \ldots = 3{,}6315$$

$$\hat{\pi} \quad (AUM = 0, ADES = 0, ANCM = 0, TAMCPDESCR = 5, PROPV = 5, QTDENCMS = 1, VVMI = 15, REGIAO = SE, SECAO = 1)$$

$$= \frac{e^{(3{,}6315)}}{1 + e^{(3{,}6315)}} = 0{,}78$$

Tabela 7. Proporção estimada de respostas conforme a unidade de medida IBGE, Pia-Produto, 2016 a 2018

	Exemplo 1			Exemplo 2			Exemplo 3			Exemplo 4			Exemplo 5		
	2016	2017	2018	2016	2017	2018	2016	2017	2018	2016	2017	2018	2016	2017	2018
Proporção estimada de respostas de acordo com a u.m. IBGE	0,78	0,79	0,80	0,36	0,46	0,46	0,49	0,51	0,60	0,67	0,75	0,70	0,79	0,80	0,87
Alteração: Unidade de Medida	Não			Sim			Sim			Não			Sim		
Alteração: Descrição	Não			Sim			Não			Sim			Sim		
Alteração: NCM	Não			Sim			Não			Sim			Sim		
Nº palavras no campo de descrição	5			5			5			5			5		
Participação na Classe	5			10			5			10			30		
Nº de códigos NCM	1			1			1			1			1		
Valor de Vendas (R$ milhões)	15			15			15			15			15		

Fonte: Pia Produto 2016 a 2018 e Prodlist 2016

Algumas covariáveis que são estatisticamente significativas podem ter esta relação com a variável resposta apenas em alguns pontos da distribuição. Por este motivo é comum que elas sejam separadas em categorias ordenáveis de forma que se consiga entender quais são os pontos onde há uma associação significativa com a variável resposta. Por exemplo, o número de palavras (TAMCPDESCR) foi testada sob as seguintes categorias[24]: (i) até 10 palavras (categoria de referência); (ii) de 11 a 15 palavras; e (iii) mais de 15 palavras. A receita líquida de vendas foi dividida em três categorias, (i) até R$ 50 mi (categoria de referência); (ii) mais R$ 50 mi e menos de R$ 100 mi; e (iii) R$ 100 mi ou mais. O número de códigos NCM para código do produto na Prodlist foi dividido da seguinte maneira: (i) até 5 NCMs por cdprod (categoria de referência); (ii) de 6 a 10 NCMs por cdprod; e (iii) 11 ou mais NCMs por cdprod. Desta forma é possível que se estabeleça uma relação associativa escalonada, entre as covariáveis, agora em formato categórico ordenável, e a variável resposta.

Os resultados apresentados na tabela 8 são similares aos da tabela 5, trocando-se apenas as variáveis numéricas por categóricas. Quando há a separação das variáveis em categorias alguns resultados ficam mais claros como o caso do número de palavras em classes. Para esta variável a categoria de referência são as descrições com até 10 palavras. No entanto, os dois níveis (11 a 15 e + 15 palavras) apresentaram níveis muito parecidos levando a conclusões parecidas. Interpreta-se, neste caso específico, que os campos de descrição com onze ou mais palavras em relação às descrições com poucas palavras têm um efeito negativo na resposta da unidade de medida igual à do IBGE. Pode-se verificar que a separação em três categorias para esta variável não trouxe benefícios para a distinção dos coeficientes estimados, pois eles proporcionam o mesmo resultado. Olhando-se para a tabela 9, que possui as razões de chance, depreende-se que os produtos cuja definição possui mais de 10 palavras em relação aos produtos cuja definição possui até 10 palavras diminui as chances dos informantes entregarem a informação na unidade de medida IBGE em torno de 20% em 2016 (100 x [RC – 1] = – 0,2).

Os produtos que possuem receita líquida de vendas na segunda categoria (R$ 50 mi – 100 mi) não apresentam resultados estatisticamente diferentes aos produtos que

.....

[24] As categorias foram criadas a partir dos quantis da distribuição do número de palavras: mínimo = 1; Q1 = 6; 50% = 9; Q3 = 13; 90% = 17; 95% = 20; 99% = 27; e máximo = 35.

estão na categoria de referência (até R$ 50 mi) no que diz respeito à entrega da informação na unidade de medida IBGE. No entanto, verifica-se que os produtos que possuem uma RLV maior do que R$ 100 mi em relação à primeira categoria apresentaram um valor positivo e significativo. Em 2016 essas empresas tinham aproximadamente 1,173 a mais de chances de entregarem a informação na unidade de medida exigida pelo IBGE do que um produto que não tinha essa característica específica naquele mesmo ano. Em 2018 as chances passaram para 1,44 vezes.

O número de códigos NCM apresentou resultado interessante, pois nas duas categorias de comparação resultados distintos foram estimados. Os produtos que possuem 6 a 10 códigos NCM para cada cdprod em relação aos produtos com até 5 códigos NCM apresentaram resultados piores em relação à entrega da informação na unidade de medida IBGE apesar de haver uma melhora ao longo dos três anos. Interessante notar que os produtos com mais de 10 códigos NCM em relação à primeira categoria apresentaram resultados significativos para 2016 e 2018. No entanto, as razões de chance mostram que essa melhora não é tão grande assim.

Tabela 8. Resultados do modelo de regressão logística com as variáveis consideradas finais, para cada ano do estudo

Variável	Nível da Variável	Par.Est. 2016	Par.Est. 2017	Par.Est. 2018
Intercepto		0,77*	0,833*	1,056*
Alteração U.M.	1= alterada	-1,305*	-1,247*	-0,987*
Alteração Descrição	1= alterada	0,214*	0,198*	0,204*
Alteração NCM	1= alterada	-0,854*	-0,495*	-0,871*
Nº palavras - classe	11-15 palavras	-0,071*	-0,07*	-0,069*
(ref=até 10 palavras)	15 ou mais palavras	-0,076*	-0,062*	-0,062*
Valor de Vendas - classe	R$ 50 mi a R$ 100 mi	0,005	-0,036	0,099*
(ref=até R$ 50 mi)	R$ 100 mi ou mais	0,077*	0,105*	0,135*
Códigos de NCM - classe	6 a 10 NCMs	-0,157*	-0,111*	-0,093*
(ref=até 5 NCMs)	10 ou mais NCMs	0,083*	0,021	0,07*
Participação Classe		0,023*	0,02*	0,042*
Região	Norte	-0,137*	-0,231*	-0,108*
(ref= Região Sudeste)	Nordeste	0,235*	0,184*	0,27*
	Sul	-0,096*	0,028	-0,113*
	Centro Oeste	0,009	0,08*	-0,012
Seção	Zero	0,817*	1,153*	1,029*
(ref=seção 1)	Dois	-0,182*	-0,219*	-0,33*
	Três	0,039	-0,172*	-0,168*

Fonte: Pesquisa Industrial Anual – Produto, 2016, 2017 e 2018.

Nota: * coeficientes estimados significativos ao nível de significância de 5%.

A análise das variáveis AA, AC, AD, AM e AN deve ser feita com bastante critério uma vez que as comparações dos seus coeficientes estimados devem ser feitas em relação à última edição da Prodlist (2013). Por exemplo, para os três anos pode-se notar que a alteração dos produtos em relação à unidade de medida (AM) não foi benéfica aos respondentes. No entanto, ao longo dos três anos, aparentemente houve um aprendizado dos informantes em relação ao preenchimento da unidade de medida uma vez que os coeficientes estimados indicaram uma crescente melhora.

A única alteração nos códigos de produtos benéfica para que obtivesse respostas na unidade de medida IBGE foi a alteração na descrição dos produtos. A intepretação formal para esta definição é a seguinte: produtos cuja descrição foi alterada entre as edições da Prodlist de 2013 e 2016 apresentaram maior probabilidade de terem tido respostas na unidade de medida requerida pela Pesquisa, em relação aos produtos em que não houve tal alteração. Os resultados para todos os anos foram bastante parecidos. Os produtos nesta condição tiveram aproximadamente 24% mais chances de terem as unidades de medida respondidas corretamente quando comparados aos produtos que não tiveram alteração na descrição da edição da Prodlist 2013.

Tabela 9. Razão de Chances referentes ao modelo apresentado na tabela 7, para cada ano do estudo

Variável	Nível da Variável	Estimativa 2016	Estimativa 2017	Estimativa 2018
Alteração U.M.	1= alterada	0,27 (0,26;0,28)	0,29 (0,28;0,3)	0,37 (0,36;0,39)
Alteração Descrição	1= alterada	1,24 (1,16;1,32)	1,22 (1,15;1,3)	1,23 (1,15;1,31)
Alteração NCM	1= alterada	0,43 (0,35;0,52)	0,61 (0,51;0,73)	0,42 (0,35;0,5)
Nº palavras	11 a 15 vs até 10 palavras	0,8 (0,78;0,83)	0,82 (0,79;0,85)	0,82 (0,79;0,85)
	+15 vs até 10 palavras	0,8 (0,76;0,84)	0,83 (0,79;0,86)	0,82 (0,79;0,86)
Classes Valor de Vendas	R$ 50 a 100 mi vs até R$ 50 mi	1,09 (1;1,19)	1,03 (0,95;1,12)	1,4 (1,29;1,52)
	R$ 100 mi + vs até R$ 50 mi	1,17 (1,08;1,27)	1,19 (1,1;1,29)	1,45 (1,33;1,57)
Nº de códigos NCM	6 a 10 NCMs vs até 5 NCMs	0,79 (0,75;0,84)	0,82 (0,78;0,86)	0,89 (0,85;0,94)
	10 + NCMs vs até 5 NCMs	1,01 (0,96;1,06)	0,93 (0,89;0,98)	1,05 (1;1,1)
Participação Classe	proporção	1,02 (1,01;1,04)	1,02 (1,01;1,03)	1,04 (1,03;1,06)
Região	N vs SE	0,88 (0,8;0,97)	0,84 (0,76;0,93)	0,93 (0,85;1,02)
	NE vs SE	1,28 (1,21;1,35)	1,28 (1,21;1,35)	1,36 (1,29;1,44)
	S vs SE	0,92 (0,89;0,95)	1,09 (1,06;1,13)	0,93 (0,9;0,96)
	CO vs SE	1,02 (0,94;1,11)	1,15 (1,06;1,25)	1,03 (0,95;1,11)
Seção	Zero vs Um	4,44 (3,91;5,04)	6,79 (5,89;7,83)	4,76 (4,12;5,5)
	Zero vs Dois	1,64 (1,58;1,69)	1,72 (1,67;1,78)	1,22 (1,18;1,26)
	Zero vs Três	2,04 (1,92;2,18)	1,8 (1,7;1,92)	1,44 (1,35;1,53)

Fonte: Pesquisa Industrial Anual - Produto, 2016, 2017 e 2018.

Nota: * coeficientes estimados significativos ao nível de significância de 5%.

Conclusões e sugestões

Ao término deste trabalho foi importante entender que as informações constantes nas bases de dados da Pesquisa Industrial Anual Produto, bem como a utilização das informações constantes da relação dos produtos dessa pesquisa – a Prodlist (incluídas aqui todos os tipos de alterações nos códigos) foi o bastante para que conclusões importantes fossem feitas para um processo futuro de crítica e imputação dos dados daquela Pesquisa.

Cabe lembrar que uma importante covariável não foi utilizada por motivos óbvios de endogeneidade. A variável em questão diz respeito à variedade de respostas que um determinado produto obteve nas pesquisas. Por exemplo, há casos onde um único produto foi respondido apenas em uma única unidade de medida, a correta. No entanto, há casos cujo número de respostas em relação à unidade de medida foi bastante variado, como é o caso dos produtos plásticos que podem chegar a nove tipos de unidades de medida diferentes. A endogeneidade vem do fato de que quanto maior for a quantidade de respostas diferentes para um determinado produto as chances do informante entregar a informação na unidade de medida correta é menor. E quanto menor for a variedade de respostas para uma unidade de medida específica, maior serão as chances de entregar a informação na unidade de medida requerida pela Pesquisa. Dessa forma, ela não foi incluída no modelo por ter uma relação direta com a variável resposta do modelo.

As variáveis relativas à alteração nos códigos dos produtos foram importantes para que se pudesse entender que apenas as alterações nas descrições foram benéficas para que os respondentes entregassem a informação da forma que a Pesquisa requereu. Seus coeficientes estimados foram positivos e significativos. Já as alterações nas unidades de medida foram as que influenciaram negativamente na entrega da resposta conforme prevista pela Pesquisa, apesar de ter havido pequena melhora ao longo dos anos. Ou seja, alterações nas unidades de medida devem de alguma forma confundir os informantes no momento do preenchimento do campo: “unidade de medida”. Esta mesma constatação pôde ser verificada com as alterações relativas aos códigos NCM. Produtos com alterações nos códigos NCM também não tiveram boas respostas para as unidades de medida.

O número de palavras no campo de descrição teve efeito importante nas respostas das unidades de medida. Ficou claro que produtos cujos campos de descrição contêm menos de 10 palavras tendem a ter melhores respostas nas unidades de medida IBGE do que campos maiores do que aquele número.

Produtos cuja receita líquida de vendas é maior que R$ 100 mi tiveram respostas melhores que aqueles com valores menores do que o mesmo corte para as unidades de medida.

A quantidade de códigos NCM revelou resultados importantes. Cdprods com mais de 10 NCMs revelaram resultados bastante parecidos com os de menos de 5 NCMs por cdprod. Enquanto estes tiveram relação positiva com as respostas na unidade de medida

IBGE, os códigos de produtos com 6 a 10 NCMs não tiveram boa resposta em relação às unidades de medida.

A RLV do produto em relação à RLV total da classe mostrou que os produtos-UL mais importantes em cada classe entregam a informação na unidade de medida do IBGE. Em média, a cada ponto percentual acrescido de RLV em uma classe econômica as chances de um produto ter sua unidade de medida respondida corretamente aumenta à taxa de 1,02 em 2016 e 2017 e 1,04 em 2018. Ou seja, como era de se esperar as unidades locais mais representativas em cada setor tendem a responder melhor a Pesquisa no que diz respeito à unidade de medida do que as menos representativas. O termo quadrático PPROV significativo mostra que há um ponto limite para o aumento dessa taxa. Existe um número limite de casos para cada classe.

Outro fato interessante a ser mencionado é que, em média, as respostas referentes à unidade de medida IBGE são melhor preenchidas pelas empresas da região Nordeste e Centro Oeste quando comparadas às respostas dadas pelas empresas da região Sudeste (região de referência).

Após a verificação das variáveis que realmente foram importantes no entendimento do problema discutido, as conclusões retiradas deste estudo poderão melhorar a forma pela qual será feita a crítica dos dados da pesquisa em suas próximas edições e, também, na revisão das próximas edições das listas de produtos, as Prodlists. A busca pelo preenchimento mais eficaz dos informantes deve-se tornar prioritária na pesquisa. Talvez a utilização de novas bases de dados, tais como a base do CEMPRE aliadas à Pesquisa Anual Industrial Empresa possam trazer informações importantes em um futuro trabalho. É possível que o emprego de um indicador de produtividade, utilizando informações de Pessoal Ocupado e Receita Bruta/Líquida, seja importante, uma vez que alguns autores apontam que a produtividade dentro do mesmo setor da economia tende a ser parecida somente nas empresas de grande porte, bem estruturadas, e diferentes nos setores de menor porte (Infante, R. et al, 2015).

Uma variável que não foi incluída na análise por não ser captada pela pesquisa é o tipo de respondente: (i) profissional externo à empresa (geralmente um contador); (ii) profissional interno da área administrativa da empresa; ou (iii) profissional interno ligado à atividade produtiva da empresa. No caso (i) a pessoa pode não ter a compreensão da atividade econômica exercida pela empresa e por este motivo não possuir as informações necessárias para responder as questões relativas à atividade produtiva. Dessa forma, a unidade de medida pode ser considerada informação secundária, pelo informante, e o seu mau preenchimento ligado a esse fato. Na próxima atualização do questionário (Módulo Informante) seria importante que fosse inserida no rol de perguntas da Pesquisa Industrial Anual – Produto.

Referências

Achen, C.H. (1985) *Proxy Variables and Incorrect Signs on Regression Coefficients*, Political Methodology Vol. 11, No. 3/4 (1985), pp. 299-316. https://www.jstor.org/stable/41289346

Agresti, A. (2007) *An Introduction to Categorical Data Analysis*, 2ª edição, John Wiley & Sons.

Allison, P. D. (2005) Fixed Effects Regression Methods for Longitudinal Data Using SAS, Cary, NC: SAS Institute.

Allison, P. D. (2006*) Fixed Effects Regression Methods In SAS*®, Paper 184-31, Universidade da Pensilvânia, disponível em: http://www2.sas.com/proceedings/sugi31/184-31.pdf

Allison, P. D. (1999) *Logistic Regression Using the SAS System*, Cary, NC: SAS Institute Inc.

Baum, Christopher F. (2006) An introduction to modern econometrics using Stata, Stata Press.

De Wall, et al. (2011), Handbook of statistical data editing and imputation, John Wiley & Sons, Statistics Netherlands.

Elkin, E. (2012) *Beyond Binary Outcomes: PROC LOGISTIC to Model Ordinal and Nominal Dependent Variables*, Paper 427-2012, ICON Late Phase & Outcomes Research, San Francisco, CA, USA. https://support.sas.com/resources/papers/proceedings12/427-2012.pdf

Greene, W. H. (2002), *Econometric Analysis*, 5ª edição. Prentice-Hall, Englewood Cliffs, NJ.

Gujarati, D.N. (2000), Econometria Básica, 3ª edição. Pearson Makron Books, São Paulo.

Horstman, J.M. (2018) *Doing More with the SGPLOT Procedure*, SESUG Paper 205

Horstman, J. M. (2018) *Getting Started with the SGPLOT Procedure*. SAS Users Group 2018

Hosmer, D.W., Lemeshow, S., and Sturdivant, R.X. (2013). *Applied Logistic Regression*, 3ª edição, John Wiley & Sons,

Infante, R, Mussi, C., Oddo, M. (2015). Por um desenvolvimento inclusivo O caso do Brasil, Cepal e Ipea.

Lista de Produtos da Indústria - Prodlist - Indústria 2016, IBGE

Osborne, J.W. (2015) *Best practices in logistic regression*, Univ. de Louisville, Sage

Schwartz, S. (2009), *Clinical Trial Reporting Using SAS/GRAPH® SG Procedures*, Paper 174, SAS Institute Inc., Cary, NC

Senaviratna, N.A.M.R et ali., Diagnosing Multicollinearity of Logistic Regression Model, Asian Journal of Probability and Statistics, Sri Lanka.

Suchower, L. J. e Copenhaver, M. D. Using the SAS system to perform Mcnemar´s test and calculate the kappa statistic for matched pairs of data. Astra merck, Inc. 725 Chesterbrook Blvd, Wayne, PA, 19087.

Squeff, Gabriel Coelho A Heterogeneidade estrutural no Brasil de 1950 a 2009/Gabriel Coelho Squeff/ Mauro Oddo Nogueira. Brasília, DF: CEPAL. Escritório no Brasil/IPEA, 2013.

Wooldridge, J. M. (2001), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

Sites acessados para pesquisa:

https://stats.idre.ucla.edu/r/dae/mixed-effects-logistic-regression
https://stats.idre.ucla.edu/other/mult-pkg/introduction-to-generalized-linear-mixed-models/
https://stats.idre.ucla.edu/sas/faq/in-proc-logistic-why-arent-the-coefficients-consistent-with-the-odds-ratios/
https://stats.idre.ucla.edu/sas/faq/how-can-i-get-adjusted-predicted-values-from-a-logistic-model-in-sas/

Anexo

Por que considerar o fator tempo na análise?

A compreensão da invariabilidade das informações no tempo pode tornar a análise mais forte no sentido de que independente do ano que se faz a análise a constância das interpretações é robusta o suficiente para se tirar conclusões que não sofram distúrbios ao longo do tempo.

Uma das formas de se constatar que houve alterações na estrutura das respostas entre dois anos quaisquer é simplesmente cruzando as duas variáveis em uma tabela 2 x 2. O nível de concordância e discordância da característica que se pretende entender pode direcionar as conclusões do estudo.

Tabela A.1. Comparação entre anos no preenchimento das unidades de medida, 2016 a 2018

		U.M. 2017			U.M. 2018		
		SIM	NÃO	Total	SIM	NÃO	Total
U.M. 2016	SIM	**38,51**	11,38	49,89	**35,72**	13,6	49,32
	NÃO	12,63	**37,48**	50,11	19,00	**31,69**	50,69
		51,14	48,86	100	54,72	45,29	100
U.M. 2017	SIM				**38,88**	11,88	50,76
	NÃO				15,69	**33,55**	49,24
					54,57	45,43	100

Fonte: Pesquisa Industrial Anual Produto, 2016, 2017 e 2018.

Comparando duas a duas, todas as combinações dos três anos, observou-se que a proporção de casos concordantes sempre esteve maior que o caso de discordâncias, sendo que a comparação dos anos de 2016 e 2018 apresentou a menor concordância, 67,41%. 2016 x 2017 - 75,99% e 2017 x 2018 - 72,43%. No entanto, para que a comparação da estrutura entre anos fique mais precisa é correto avaliar somente os casos discordantes.

Com isso, verificando-se a diferença nas proporções dos casos de Unidades Iguais enxerga-se melhor a evolução nas respostas que realmente interessam. Em 2016 49,9% responderam na Unidade de Medida indicada pelo IBGE, em 2018 esse percentual passa a ser de 54,72%. Apesar das proporções poderem ser próximas, porém crescentes, é interessante que essa diferença seja relevante do ponto de vista estatístico. Pelo fato de não serem consideradas amostras independentes, uma vez que grande parte das empresas é a mesma entre os anos da pesquisa, deve-se utilizar um teste estatístico que leve em consideração esta característica, o teste estatístico McNemar[25]. Desconsiderando-se os casos concordantes, a cálculo do teste estatístico dá-se da seguinte forma: um teste qui-quadrado, com um grau de liberdade, $((13811-9885)^2/(13811+9885) = 650,47)$ mostra que o p-valor é menor que 0,00001. Desta forma, pode-

[25] O teste de McNemar é usado quando se compara valores que vêm de porções de informação dependentes. Possui as características e interpretações similares ao teste qui-quadrado.

se concluir que o crescimento dos casos de respostas esperadas é estatisticamente significativo. A partir desta informação deve-se levar em consideração que o tempo é fator importante na compreensão da análise do mau preenchimento da variável unidade medida da Pia-Produto.

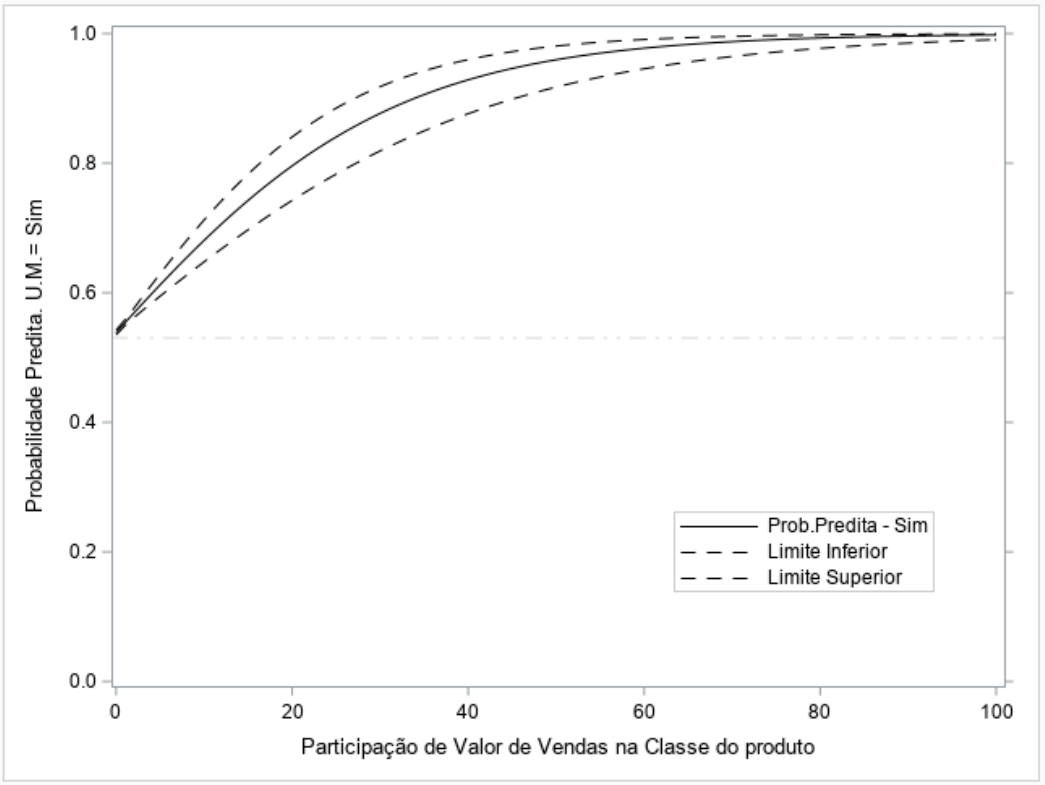
Segundo os dados da tabela 2016 x 2018 pode-se dizer que as chances de um informante passarem a responder na unidade de medida foi de 1.24.

Como é o propósito deste estudo de também estudar como seria análise caso levasse em consideração o efeito fixo (o erro fosse correlacionado com as covariáveis) a proporção de acerto seria bem maior, 13.881/9.885=1,40.

Termo Quadrático - PROPV

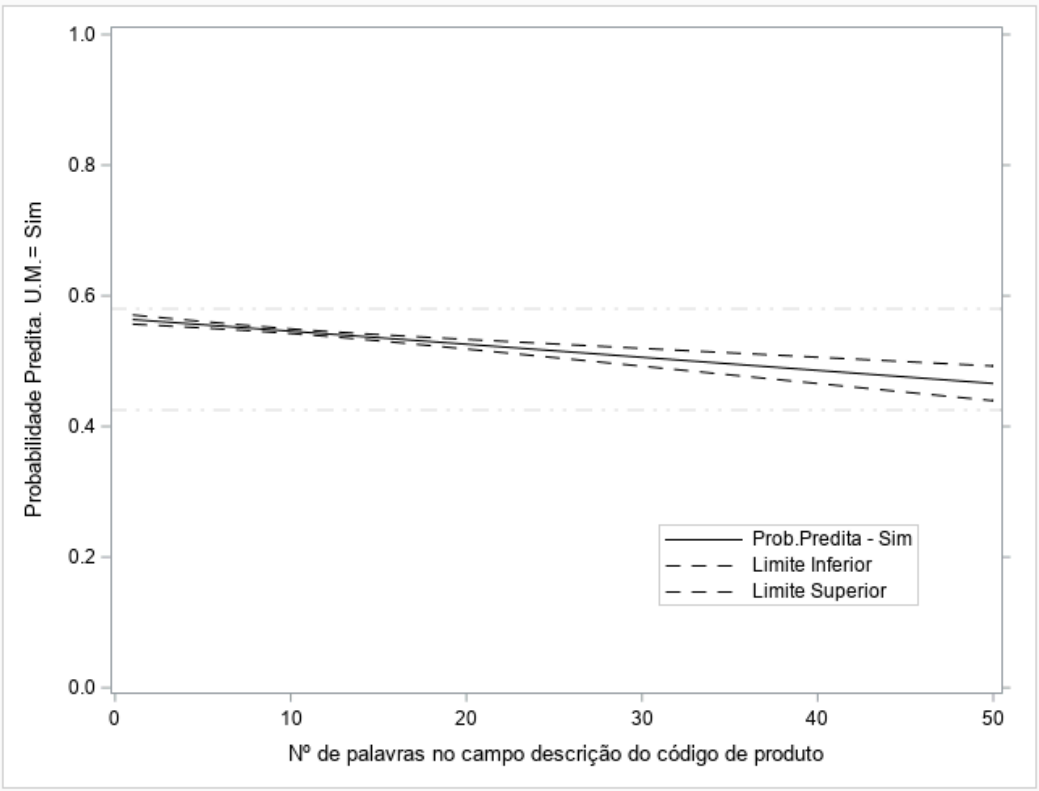
No modelo apresentado na tabela 5 foi incluído o termo quadrático para a variável Participação Classe. Ele é utilizado para que se verifique se a variável que compõe o modelo pode ter um efeito não linear na variável resposta. Isto acontece com a PROPV. Há um determinado ponto cuja uma unidade adicional de Participação da Classe não interfere na probabilidade de resposta na unidade de medida IBGE pelo informante. Pelo fato desse termo ser significativo é importante que componha o modelo final. O gráfico A.1, indica que há um ponto em que a taxa de crescimento começa a ser cada vez menor.

Gráfico A.1. Probabilidades preditas para resposta da unidade de medida ter sido respondida igual à requerida pelo IBGE no caso da Participação do Valor de Vendas na Classe do Produto, Pia-Produto – 2018



Fonte: Pia-Produto, 2016, 2017 e 2018.

Gráfico A.2. Probabilidades preditas para resposta da unidade de medida ter sido respondida igual à requerida pelo IBGE no caso do número de palavras no campo de descrição do código do produto, Pia-Produto – 2018.



Fonte: Pia-Produto, 2016, 2017 e 2018.

Tabela A.2. Correlação de Pearson entre as variáveis disponíveis para análise, Pia-Produto – 2016 a 2018

	Códigos NCM			Log (Valor de Vendas)			Nº de palavras			Share RLV (Classe)			Valor de Vendas			Log (Códigos NCM)		
Códigos NCM	2016	2017	2018	2016	2017	2018	2016	2017	2018	2016	2017	2018	2016	2017	2018	2016	2017	2018
2016	1			-0,017			0,112			0,001			-0,008			0,801		
2017		1			-0,012			0,124			0,002			-0,002			0,806	
2018			1			-0,006			0,120			0,002			-0,004			0,802
Log (Valor de Vendas)																		
2016				1,000			0,003			0,291			0,241			-0,035		
2017					1,000			0,001			0,274			0,095			-0,028	
2018						1,000			-0,008			0,285			0,225			-0,028
Nº de palavras																		
2016							1,000			0,005			-0,001			0,190		
2017								1,000			0,007			-0,001			0,196	
2018									1,000			0,007			0,004			0,197
Share RLV (Classe)																		
2016										1,000			0,423			0,002		
2017											1,000			0,282			0,003	
2018												1,000			0,395			0,004
Valor de Vendas																		
2016													1,000			-0,010		
2017														1,000			0,000	
2018															1,000			-0,006
Log (Códigos NCM)																		
2016																1		
2017																	1	
2018																		1

Fonte: PIA Produto - 2016, 2017 e 2018.

Nota: Valores hachurados são não significativos ao nível de significância de 10%.

Tabela A.3. Alterações nos códigos dos produtos: Proporção de produtos alterados nas amostras da Pia-Produto, 2016 a 2018

Categorias de Alteração de cdprods	AA	AC	AD	AM	AN	AG	DG	E	I
2016	-	0,6%	6,8%	33,5%	0,6%	1,3%	0,7%	0,02%	0,3%
2017	-	0,5%	6,7%	33,7%	0,7%	1,2%	0,8%	0,01%	0,3%
2018	-	0,5%	6,8%	34,0%	0,7%	1,2%	0,7%	0,01%	0,3%

Alterações: (AA) Atividade/cod Prodlist; (AC) Conteúdo; (AD) Descrição; (AM) UM; (AN) NCM; (E) Exclusão; e (I) Inclusão

Tabela A.4. Alterações nos códigos dos produtos (combinações de agregações): Número de produtos alterados nas amostras da Pia-Produto, 2016 a 2018

Categorias de Alteração de Codprods	AA•AM	AC•AM	AC•AN•AM	AC•AN	AD•AM	AD•AN•AM	AD•AN	AN•AM
2016	0,07%	0,15%	0,05%	0,16%	4,88%	0,10%	0,24%	0,18%
2017	0,06%	0,16%	0,05%	0,14%	4,81%	0,08%	0,22%	0,20%
2018	0,06%	0,18%	0,05%	0,13%	4,97%	0,08%	0,26%	0,19%

Alterações: (AA) Atividade/cod Prodlist; (AC) Conteúdo; (AD) Descrição; (AM) UM; (AN) NCM; (E) Exclusão; e (I) Inclusão

Se o assunto é **Brasil**,
procure o **IBGE**.

 /ibgecomunica  /ibgeoficial  /ibgeoficial  /ibgeoficial

www.ibge.gov.br 0800 721 8181

ISBN 978-65-87201-55-9