

Importância das distribuidoras para o sucesso de público dos filmes de ficção nacionais entre 2001 e 2013

25 de fevereiro de 2015

1 Objetivo

Discriminar e quantificar o impacto das distribuidoras sobre o público nos filmes ficcionais brasileiros no mercado doméstico.

2 Material e Métodos

Os métodos utilizados no estudo estão resumidos na Figura ???. Descrições detalhadas de cada etapa estão apresentadas abaixo.

2.1 Coleta de dados relacionados aos filmes

Dados referentes aos filmes lançados no período de 2001 à 2013 foram obtidos em bancos de dados de organizações governamentais (ANCINE; Ministério da Justiça), sites relacionados à cinema (IMDB, Rotten Tomatoes, adoro-cinema, etc) e em periódicos (Veja, Folha de São Paulo, Estadão, etc). A tabela presente no arquivo *tabela1A.csv* (anexado) contém todos os dados levantados e as respectivas fontes das informações.

2.2 Remoção de filmes de acordo com filtros

Filmes foram removidos se:

1. Eram documentários.
2. Não possuíam informação de variáveis usadas no modelo.
3. Eram relançamentos.

O número de filmes removidos e as causas estão na Tabela ???. Dos 1020 filmes 598 foram removidos.

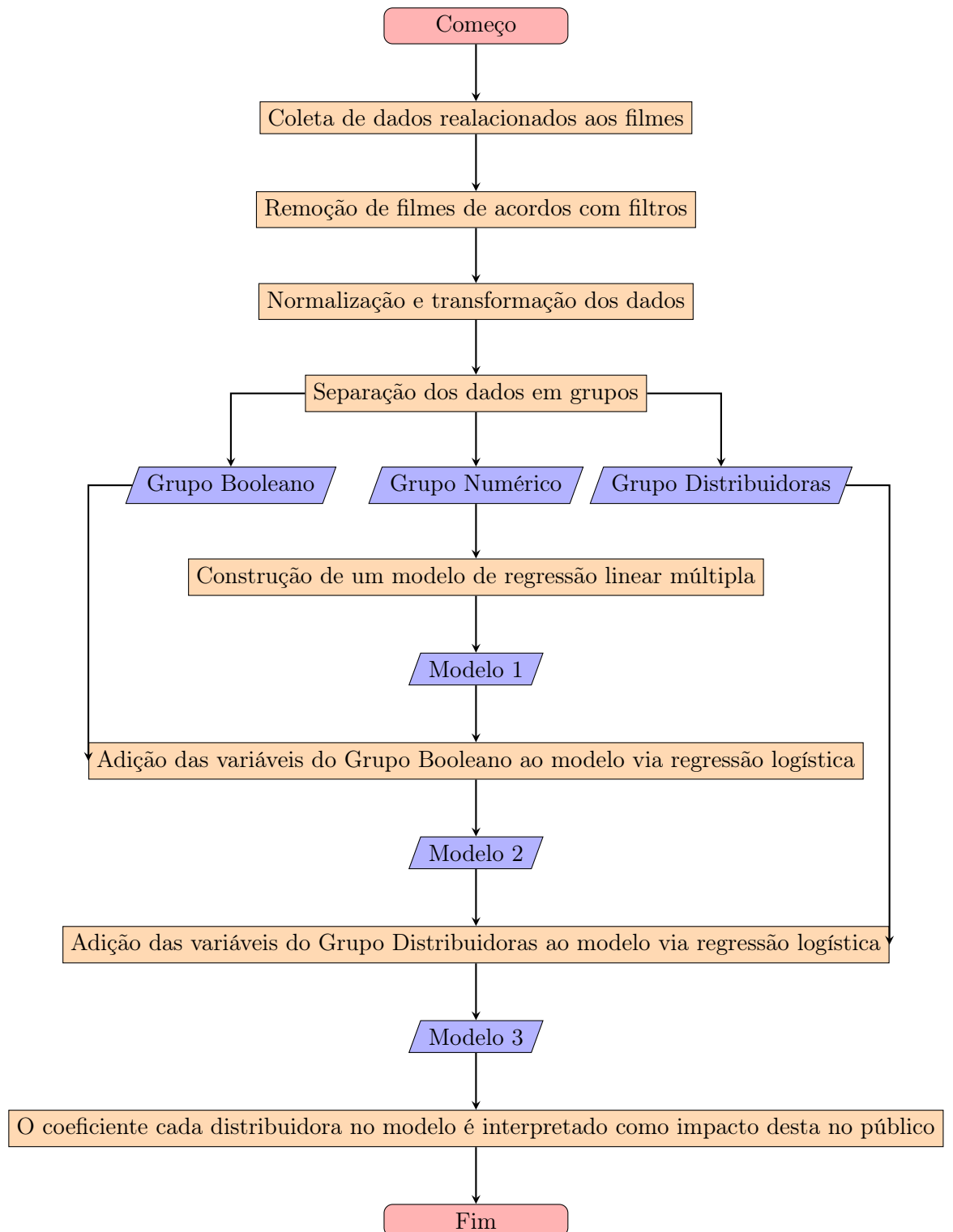


Figura 1: Fluxograma resumido dos métodos utilizados no estudo

Motivo da remoção	Quantidade de filmes
Eram Documentários	373
Ausência do valor captado junto a ANCINE	141
Ausência de avaliação qualitativa do filme	1
Ausência do público do filme	2
Eram relançamento	81

Tabela 1: Quantidade de filmes removidos do estudo de acordo com as causas.

2.3 Normalização e transformação dos dados

Os valores captados junto à ANCINE foram ajustado segundo a inflação para valores de dezembro de 2013 utilizando os índices fornecidos pelo Instituto de Pesquisa Econômica Aplicada (IPEA).

As notas qualitativas dos filmes segundo periódicos ou *site* foram normalizadas e arredondadas para o número intergral mais próximo no intervalo de 1 à 3. Isso porque as fontes destas notas são múltiplas, com diferentes intervalos de pontuações e, presumidamente, possuem heterogenidade nos critérios de avaliação.

Campos multivalorados foram transformados em variáveis booleanas de acordo com a presença ou ausência da variável nos filmes (ver caixa *Campos multivalorados à campos booleanos*). Para evitar aumentar o número de comparações e também para incrementar a robustez dos testes somente se analisou variáveis booleanas contidas em mais de 8 filmes. A lista das variáveis booleanas após o filtro e o número de filmes em que elas estão presentes se encontram na Tabela ??.

Campos multivalorados à campos booleanos

Para exemplificar processo vamos usar o campo *atores* dos filmes *Tropa de Elite 2* e *O Caminho das Nuvens*. Os Atores principais de cada filme estão listados abaixo:

Tropa de Elite 2

- Wagner Moura
- Maria Ribeiro

O Caminho das Nuvens

- Cláudia Abreu
- Wagner Moura
- Ravi Ramos Lacerda

A presença ou não de cada ator (Wagner Moura, Maria Ribeiro, Cláudia Abreu e Ravi Ramos Lacerda) será representada em um campo. Nesse o valor *1* será atribuído aos filmes que apresentam o ator e o valor *0* aos filmes que não o apresentam.

Dessa forma ao invés da notação:

Filme	Atores
Tropa de Elite 2	W. Moura / M. Ribeiro
O Caminho das Nuvens	C. Abreu / W. Moura / R. Lacerda

Teremos a notação:

Filme	W. Moura	M. Ribeiro	C. Abreu	R. Lacerda
Tropa de Elite 2	1	1	0	0
O Caminho das Nuvens	1	0	1	1

Essa transformação foi feita em todos os campos multivalorados (*Diretores*; *Atores*; *Gêneros*; *Distribuidoras*). As colunas com soma menor que oito foram removidas. Ou seja, só foram considerados atores diretores, gêneros ou distribuidoras com presença em pelo menos oito filmes.

Uma decorrência desse filtro é que somente empresas que distribuíram mais de 8 filmes (*distribuidoras atuantes* – *DA*) seriam analisadas. Porém, notou-se que todas as *DA* que significativamente impactavam o público o faziam de forma positiva, como será demonstrado abaixo nos resultados. Se conjecturou que como as *DA* tem grande presença no mercado é esperado

Tipo	Variável	Quantidade de filmes
Diretores	Daniel Filho	8
	Moacyr Goes	9
Atores	João Miguel	8
	Mariana Ximenes	8
	Marco Ricca	8
	Dira Paes	9
	Rodrigo Santoro	9
	Fábio Assunção	9
	Lázaro Ramos	9
	Leonardo Medeiros	9
	Caco Ciocler	10
	Paulo José	10
	Wagner Moura	11
	Caio Blat	11
	Selton Mello	14
Gêneros	Infantil	11
	Animação	13
	Suspense	13
	Ação	16
	Aventura	25
	Romance	35
	Comédia	97
	Drama	253
Distribuidoras	Estação	10
	Paramount	11
	Paris	11
	Providence	18
	Imovision	20
	RioFilme	25
	Europa	26
	Fox	27
	Warner Bros.	29
	Downtown	29
	Imagem	33
	Columbia	52
	DO*	131

* *Distribuidoras ocasionais*: Categoria atrelada aos filmes que não foram distribuídos por nenhuma das outras distribuidoras presentes.

Tabela 2: Variáveis booleanas consideradas no estudo

que elas tenham maior facilidade de distribuir e promover seus filmes do que as demais, o que explicaria o resultado. Para mensurar o impacto do filmes não serem distribuídos por uma *DA* e para avaliar estas distribuidoras considerando seus tamanhos se criou a classe de *distribuidoras ocasionais (DO)*. A classe *DO* foi atribuída a todos os filmes que não foram distribuídos por nenhuma das *DA*.

2.4 Separação dos dados em grupos

Os dados referentes aos filmes foram divididos em três grupos (*Grupo Numérico*; *Grupo Booleano*; *Grupo Distribuidoras*). Essa separação foi feita pois os dados de cada um dos grupos foi adicionado ao modelo em etapas (Figura ??). Os grupos estão descritos abaixo:

2.4.1 Grupo Numérico

Log(captado)

Logaritmo natural dos valores captados junto à ANCINE pelo filme ajustados segundo a inflação para valores de dezembro de 2013 utilizando os índices fornecidos pelo IPEA.

Nota

Nota qualitativa do filme segundo sites especializados em cinema ou periódicos normalizadas e arredondadas para o número intergral mais próximo no intervalo de 1 à 3.

IBOPE

Caso o filme seja baseado em um programa de TV ou tenha no título o nome de uma personalidade que apresenta um programa de TV (eg: *Casseta & Planeta: A Taça do Mundo É Nossa*; *Xuxa e os Duendes*; *Eliana e os Golfinhos*) se considerou a pontuação do programa na cidade de São Paulo do Instituto Brasileiro de Opinião Pública e Estatística (IBOPE) na data publicamente disponível mais próxima ao lançamento.

Log(Público anterior)

Caso o filme seja uma sequência (eg: *Tropa de Elite 2*, *Boleiros 2*), se considerou o público do filme anterior.

2.4.2 Grupo Booleano

Grupo formado pelas variáveis booleanas dos tipos *Diretores*, *Atores* e *Gêneros* (Tabela ??).

2.4.3 Grupo Distribuidoras

Grupo formado pelas variáveis booleanas do tipo *Distribuidora* (Tabela ??).

2.5 Construção do modelo

A construção de um modelo para a previsão do logaritmo natural do público dos filmes foi feito em três etapas. Na primeira etapa se criou a base do modelo com variáveis declaradas *a priori* importante para a previsão. Nas demais etapas se incrementou o modelo com as variáveis booleanas. Abaixo a descrição dos modelos produzidos em cada etapa:

Modelo 1

Modelo de regressão linear múltipla formado por todas variáveis do *Grupo Numérico*.

Modelo 2

Se incrementou o *Modelo 1* com as variáveis do *Grupo Booleano* usando regressão logística *stepwise forward* utilizando o Critério de Informação Bayesiano (*BIC*) como filtro através da função *StepAIC* do pacote *MASS* em *R*. Se restringiu a regressão logística para que as variáveis do *Modelo 1* não fossem retiradas.

Modelo 3

Se incrementou o *modelo 2* com as variáveis do *Grupo Distribuidoras* usando o mesmo método e critérios utilizados para construir o *Modelo 2* (inclusive restringindo para que variáveis já presentes no modelo não fossem retiradas).

Modelo 3'

Assim, como no *modelo 3* se incrementou o *modelo 2* com as variáveis do *Grupo Distribuidoras*, porém antes desse procedimento se adicionou ao modelo a variável *DO*.

2.6 Importância das distribuidoras

O coeficiente estimado de cada distribuidora no *Modelo 3* foi usado como estimativa do impacto da distribuidora (ou da variável *DO*) no público do filme.

3 Resultados

Os resultados dos modelos estão na Tabela ?? e a visualização dos impactos das distribuidoras estão apresentados das Figura ?? e na Figura ?. Abaixo a descrição dos resultados em cada um dos modelos produzidos.

3.1 Modelo 1

Todas as variáveis independentes deste modelo apresentaram correlação significativa e positiva com a variável dependente. Essa correlação está de acordo com as justificativas que nos levaram a incluir *a priori* esses campos.

<i>Variável dependente: Log(Público)</i>				
	Modelo 1	Modelo 2	Modelo 3	Modelo 3'
Constante	-9.725*** (1.231)	-7.530*** (1.178)	-4.196*** (1.238)	-4.046*** (1.246)
Log(captado)	1.278*** (0.083)	1.136*** (0.078)	0.872*** (0.085)	0.947*** (0.080)
Nota	0.717*** (0.153)	0.887*** (0.145)	0.860*** (0.138)	0.783*** (0.139)
IBOPE	0.160*** (0.031)	0.130*** (0.028)	0.130*** (0.028)	0.116*** (0.027)
Log(Público anterior)	0.121*** (0.047)	0.094** (0.043)	0.090** (0.041)	0.097** (0.041)
Gêneros				
Drama		-1.010*** (0.224)	-0.869*** (0.215)	-0.910*** (0.216)
Suspense		-1.361*** (0.494)	-1.544*** (0.471)	-1.426*** (0.471)
Comédia		0.602** (0.250)	0.374 (0.240)	0.362 (0.242)
Diretores				
Daniel Filho		1.714*** (0.616)	1.250** (0.595)	1.482** (0.589)
Moacyr Goes		1.477** (0.581)	1.487*** (0.571)	1.407** (0.554)
Atores				
Wagner Moura		1.325** (0.531)	1.225** (0.527)	1.239** (0.507)
Distribuidoras				
Paris			2.077*** (0.505)	
Fox			1.761*** (0.356)	
Downtown			1.564*** (0.332)	
Columbia			1.096*** (0.284)	
Paramount			1.497*** (0.523)	
Imagem			0.942*** (0.307)	
Warner Bros.			0.899** (0.357)	
DO				-1.211*** (0.195)
RioFilme				-1.190*** (0.349)
Providence				-1.128*** (0.413)
Observações	422	422	422	422
R ²	0.446	0.545	0.601	0.589
R ² ajustado	0.441	0.534	0.584	0.576
Erro padrão residual	1.853 (df = 417)	1.692 (df = 411)	1.599 (df = 404)	1.614 (df = 408)
F	84.063*** (df = 4; 417)	49.297*** (df = 10; 411)	35.744*** (df = 17; 404)	44.997*** (df = 13; 408)

Nota:

*p<0.1; **p<0.05; ***p<0.01

Tabela 3: Resultados das regressões lineares. Na parte de cima da tabela estão expressos os coeficientes das variáveis com o erro padrão entre parênteses.

3.2 Modelo 2

A adição das variáveis booleanas (com exceção da classe *Distribuidoras*) se deu através de regressão logística *stepwise forward* utilizando *BIC* como filtro. A formulação de modelos utilizando regressões logísticas possuem uma tendência de superajuste, ainda que se tenha tentando balancear isto com a utilização de um critério conservador (*BIC*) e fixando variáveis independentes ao modelo (as variáveis presentes no *Modelo 1*). Contudo se salienta que o *Modelo 2* não possui uma função preditiva e sim de selecionar as covariáveis presentes no *Modelo 3*, que por sua vez será utilizado somente como forma de mensurar o impacto das distribuidoras e não terá função preditiva.

Nota-se que, como apontado em publicações jornalísticas, que a variável *Comédia* aparece como tendo impacto significativamente positivo, enquanto os outros dois gêneros adicionados ao modelo (*Drama* e *Suspense*) apresentam impacto negativo. Os dois diretores analisados (Daniel Filho e Moacyr Goes) apresentaram associação significativa e positiva com a variável dependente. Importante notar que ambos os diretores estão atrelados à diversas produções ditas globais (eg: *Trair e coçar é só começar*; *Xuxa e o tesouro da cidade perdida*; *Xuxa abracadabra*; *Xuxinha e Guto contra os monstros do espaço*; *Se Eu Fosse Você*; *Se Eu Fosse Você 2*; *Primo Basílio*). Somente um ator demonstrou impacto significativo (*Wagner Moura*) e este foi positivo.

3.3 Modelo 3

Tendo selecionado as covariáveis (variáveis presentes no *Modelo 2*) para o estudo das variáveis dependentes de interesse um terceiro modelo foi criado adicionado as variáveis da classe *Distribuidoras* através também de regressão logística *stepwise forward* utilizando *BIC* como filtro. Para a construção desse modelo se considerou as 12 distribuidoras que apresentavam mais de oito filmes no período (tabela ??, classe *Distribuidoras*). Destas 12 distribuidoras 7 apresentaram um impacto significativo na variável dependente, todos positivos (Tabela ?? e Figura ??).

3.4 Modelo 3'

Se conjecturou que a razão pela qual todos os impactos das variáveis da tipo *Distribuidoras* serem positivos era devido ao viés de terem sido consideradas somente empresas com mais de 10 filmes no período, ou seja, distribuidoras que tem presença constante no mercado, logo com possíveis vantagens competitivas. Dentro do universo de filmes analisados essas empresas representa somente 11% das distribuidoras, porém estão presentes em 69% dos filmes. Como forma de inspecionar a diferença entre estas empresas e também para mensurar a importâncias delas considerando que fazem parte de um subgrupo específico se criou a classe *DO* que foi atribuída à filmes distribuídos somente por empresa(s) que distribuíram menos de 10 filmes no período.

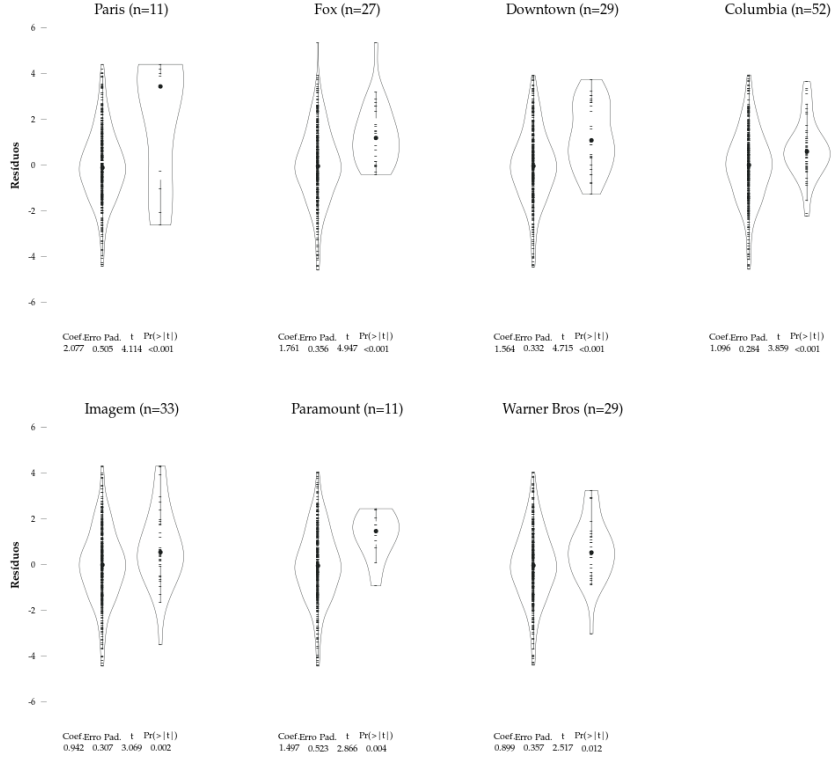


Figura 2: *Violin plot* dos resíduos de todos os filmes e dos filmes das distribuidoras com impacto significativo no *Modelo 3*. Cada gráfico representa os resíduos do *Modelo 3* desconsiderando o impacto da distribuidora em questão.

Após contemplar o *Modelo 2* com a variável *DO*, seuiu-se a adição das distribuidoras através de regressão logística, tal qual no *Modelo 3*. A variável *DO*, como esperado, teve impacto significativo e e negativo na variável dependente. Ou seja, a distribuição por empresas não atuantes (menos de 10 filmes) apresenta relação com um menor público. Além disso dentro dos parâmetros desse modelo, duas distribuidoras (RioFilme e Providence) demonstraram um impacto negativo significativo em relação à outras *DA*.

4 Conclusões

Das 12 distribuidoras analisadas, 7 demonstraram um impacto significativo e positivo sobre o público durante o período examinado. (Tabela ?? e Figura ??). O incremento esperado no público devido à presença das distribuidora variou de 207% à 89%. Uma vez estabelecido essa importância se investigou a alteração no público devido ao fato de filmes não serem distribuídos por

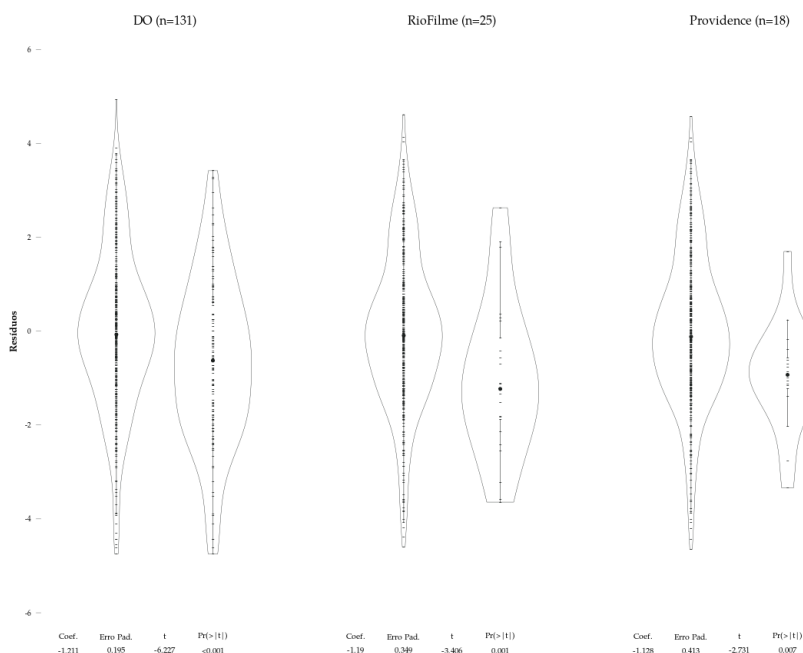


Figura 3: *Violin plot* dos resíduos de todos os filmes e dos filmes das distribuidoras com impacto significativo no *Modelo 3'*. Cada gráfico representa os resíduos do *Modelo 3'* desconsiderando o impacto da distribuidora em questão.

empresas atuantes (distribuidoras com 10 ou mais filmes no período). Essa segunda análise demonstrou que filmes que contam com pelo menos umas das duas 12 distribuidoras atuantes tendem a ter um ganho de público de 121%. Importante ressaltar que a perda de público devido ao filme ser distribuído por uma empresa não atuante não deve ser constante entre as distintas distribuidoras que compõem o grupo, porém não foi possível mensurar o impacto individual devido ao reduzido número de filmes atrelado a cada uma das empresas no período. Outro resultado dessa segunda análise é a presença de duas distribuidoras (RioFilmes e Providence) do grupo de distribuidoras atuantes, que aparecem distoante do resto do conjunto ao impactarem significativamente menos do que as outras distribuidoras atuantes (Tabela ?? e Figura ??).