

# Análise de Dados

Alessander Osorio

07/09/2023

## Sumarizações iniciais

A presente análise diz respeito aos dados oriundos do processo de validação da taxonomia proposta segundo a ontologia dela derivada. Tal ontologia, composta por 89 descritores divididos em 10 categorias, foi pesquisada no texto de 2508 artigos de 68 periódicos da ACM, totalizando 124601 observações. Estas foram submetidas ao algoritmo semisupervisionado de classificação k-nn para verificar sua adequação aos conceitos referentes a cada descritor levantados pela revisão sistemática de literatura.

Ao final, foram considerados aceitos como dentro do contexto do conceito a que se refere, 91419 observações representando 73,37% do total. Em contraposição, dentro dos mesmos padrões, 33182 observações não foram aceitas, as quais representam 26,63% do total, conforme a Tabela 1.

As Tabelas 2 e 3, mostram os totais de ocorrências por Categoria e Descritor respectivamente. Note que dado o número total de observações, alguns descritores percentualmente apresentam valores não passíveis de representação matemática com o número de casas decimais utilizadas. Todavia não é possível desconsiderar tais valores pois estes compõem o total da categoria

A variável de desfecho analisada é de natureza categorica, tendo o valor “Sim” para os aceitos, e “Não” para os não aceitos. Os valores apresentados aqui, representam a contagem de cada uma delas individualizada. As categorias e seus descritores serão analisados individualmente sem procurar estabelecer uma relação entre elas. Os histogramas por Categoria e Descritor mostram que os dados não seguem a distribuição normal, sendo o teste do Qui-Quadrado o recomendado para verificação estatística dos valores desta análise.

Table 1: Totais Gerais Aceite

Aceite	n	%
Não	33182	27
Sim	91419	73

Table 2: Totais por Categoria dos Descritores

Categoria	n	%
Ambiente	11693	9,4
Dados	5576	4,5
Estratégia de Validação/Coleta de Dados	26014	20,9
Método	40292	32,3
Metodologia	15261	12,2
Natureza	7380	5,9
Propósito paradigma	2776	2,2
Prova Científica Secundária	6977	5,6
Prova Científica	1234	1,0

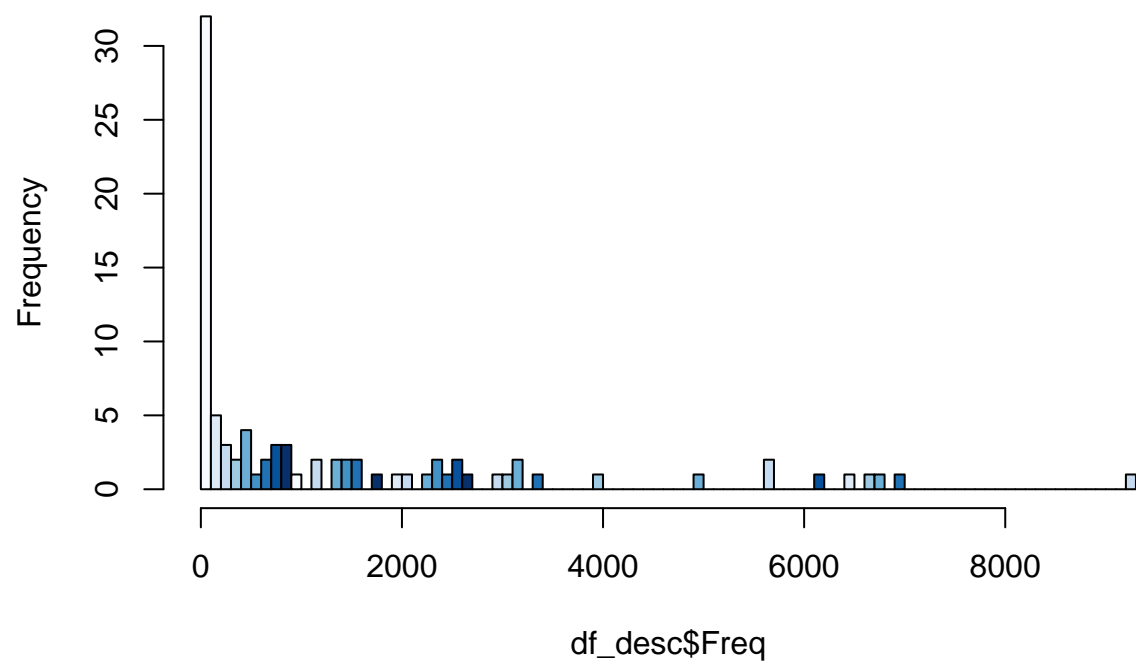
Categoria	n	%
Tipo de Validação do Resultado	7398	5,9

Table 3: Totais por Descritores

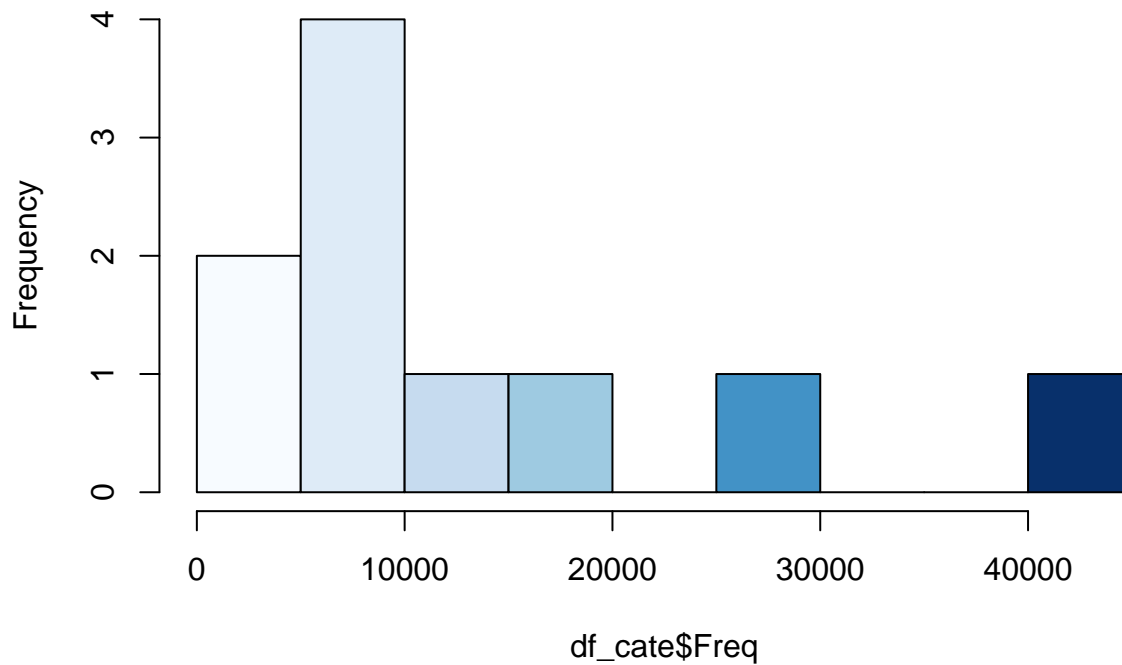
Descritor	n	%
abduction	39	0,0
action research	72	0,1
analytical	832	0,7
applied	6416	5,1
behavioral	2202	1,8
benchmark	2957	2,4
case-control	6	0,0
case study	1395	1,1
causal	1589	1,3
cause-and-effect	8	0,0
community research	10	0,0
comparative	703	0,6
concept	2531	2,0
conceptual	1171	0,9
conceptual analysis	6	0,0
constructive	401	0,3
control model	30	0,0
controlled model	1	0,0
cost-effectiveness	86	0,1
critique	400	0,3
cross-section	46	0,0
cross-sectional	75	0,1
deduction	64	0,1
descriptive	765	0,6
design science	37	0,0
doctrinal	1	0,0
dynamic analysis	273	0,2
effectiveness	2499	2,0
efficiency	3194	2,6
empiric	14	0,0
engineering	5658	4,5
evaluation	9298	7,5
evaluative	62	0,0
experiment	3978	3,2
experimental	4924	4,0
exploratory	715	0,6
formal	3383	2,7
grounded theory	237	0,2
hermeneutics	2	0,0
historic	131	0,1
historical	1365	1,1
in vitro	19	0,0
in vivo	47	0,0
induction	1131	0,9
instrument application	1	0,0

Descritor	n	%
interpretive	137	0,1
interview	2073	1,7
interviews	2306	1,9
investigate	2397	1,9
investigative	98	0,1
longitudinal	536	0,4
math	497	0,4
math test	17	0,0
meta-analysis	150	0,1
modeling	5681	4,6
normative	432	0,3
observational	483	0,4
operationalization	73	0,1
phenomenology	29	0,0
pilot	870	0,7
pilot study	179	0,1
problem-oriented	1	0,0
problem-solving	234	0,2
process analysis	4	0,0
processual	3	0,0
proof	6969	5,6
protocol analysis	3	0,0
prototype	1452	1,2
pure	964	0,8
qualitative	3153	2,5
quantitative	1762	1,4
real	6799	5,5
real-world	3028	2,4
real world	304	0,2
rebuttal	8	0,0
replication	632	0,5
scientific	1916	1,5
secondary data	18	0,0
simulated	1496	1,2
simulation	6111	4,9
speech analysis	4	0,0
standards	863	0,7
static analysis	650	0,5
survey	6692	5,4
systematic literature review	135	0,1
theoretical	2502	2,0
usability	1546	1,2
verticals	4	0,0
workload	2616	2,1

**Histograma segundo ocorrências por Descritor**



## Histograma segundo ocorrências por Categoria



## Teste Global dos aceitos/não aceitos

Esta seção mostra a aplicação dos testes estatísticos Qui-Quadrado e Teste de Proporcionalidade para toda a população das observações. O p-valor encontrado em ambos (0.0000000000000001) é inferior a 0,01 ( $p < 0,01$ ), o que rejeita a Hipótese Nula e aceita a Hipótese Alternativa para os dados analisados. Individualmente foi aplicado para cada categoria obtendo-se resultado semelhante.

Chi-squared test for given probabilities

```
data:  tabela_contingencia
X-squared = 27219, df = 1, p-value <2e-16
```

```
   Não   Sim
33182 91419
```

```
[1] "Categoria: Ambiente qui-quadrado p-valor: 0"
[1] "Categoria: Dados qui-quadrado p-valor: 0"
[1] "Categoria: Estratégia de Validação/Coleta de Dados qui-quadrado p-valor: 0"
[1] "Categoria: Método qui-quadrado p-valor: 0"
[1] "Categoria: Metodologia qui-quadrado p-valor: 0"
[1] "Categoria: Natureza qui-quadrado p-valor: 0"
[1] "Categoria: Propósito paradigma qui-quadrado p-valor: 0,14"
```

- [1] "Categoria: Prova Científica Secundária qui-quadrado p-valor: 0"
- [1] "Categoria: Prova Científica qui-quadrado p-valor: 0"
- [1] "Categoria: Tipo de Validação do Resultado qui-quadrado p-valor: 0"

## Teste Individual dos aceitos/não aceitos e descritor individualmente

- [1] "Descritor abduction qui-quadrado p-valor: 0,00023"
- [1] "Descritor action research qui-quadrado p-valor: 0,00041"
- [1] "Descritor analytical qui-quadrado p-valor: 0"
- [1] "Descritor applied qui-quadrado p-valor: 0"
- [1] "Descritor behavioral qui-quadrado p-valor: 0,061"
- [1] "Descritor benchmark qui-quadrado p-valor: 0"
- [1] "Descritor case-control qui-quadrado p-valor: Sem ocorrência de uma das categorias"
- [1] "Descritor case study qui-quadrado p-valor: 0"
- [1] "Descritor causal qui-quadrado p-valor: 0"
- [1] "Descritor cause-and-effect qui-quadrado p-valor: Sem ocorrência de uma das categorias"
- [1] "Descritor community research qui-quadrado p-valor: 0,53"
- [1] "Descritor comparative qui-quadrado p-valor: 0"
- [1] "Descritor concept qui-quadrado p-valor: 0"
- [1] "Descritor conceptual qui-quadrado p-valor: 0"
- [1] "Descritor conceptual analysis qui-quadrado p-valor: Sem ocorrência de uma das categorias"
- [1] "Descritor constructive qui-quadrado p-valor: 0"
- [1] "Descritor control model qui-quadrado p-valor: 0,00026"
- [1] "Descritor controlled model qui-quadrado p-valor: Sem ocorrência de uma das categorias"
- [1] "Descritor cost-effectiveness qui-quadrado p-valor: 0"
- [1] "Descritor critique qui-quadrado p-valor: 0"
- [1] "Descritor cross-section qui-quadrado p-valor: 0,24"
- [1] "Descritor cross-sectional qui-quadrado p-valor: 0"
- [1] "Descritor deduction qui-quadrado p-valor: 0,62"
- [1] "Descritor descriptive qui-quadrado p-valor: 0"
- [1] "Descritor design science qui-quadrado p-valor: 0,0052"
- [1] "Descritor doctrinal qui-quadrado p-valor: Sem ocorrência de uma das categorias"
- [1] "Descritor dynamic analysis qui-quadrado p-valor: 0,00005"
- [1] "Descritor effectiveness qui-quadrado p-valor: 0"
- [1] "Descritor efficiency qui-quadrado p-valor: 0"
- [1] "Descritor empiric qui-quadrado p-valor: 0,033"
- [1] "Descritor engineering qui-quadrado p-valor: 0"
- [1] "Descritor evaluation qui-quadrado p-valor: 0"
- [1] "Descritor evaluative qui-quadrado p-valor: 0,13"
- [1] "Descritor experiment qui-quadrado p-valor: 0"
- [1] "Descritor experimental qui-quadrado p-valor: 0"
- [1] "Descritor exploratory qui-quadrado p-valor: 0"
- [1] "Descritor formal qui-quadrado p-valor: 0"
- [1] "Descritor grounded theory qui-quadrado p-valor: 0"
- [1] "Descritor hermeneutics qui-quadrado p-valor: 1"
- [1] "Descritor historic qui-quadrado p-valor: 0"
- [1] "Descritor historical qui-quadrado p-valor: 0"
- [1] "Descritor in vitro qui-quadrado p-valor: 0,49"
- [1] "Descritor in vivo qui-quadrado p-valor: 0,013"
- [1] "Descritor induction qui-quadrado p-valor: 0"
- [1] "Descritor instrument application qui-quadrado p-valor: Sem ocorrência de uma das categorias"
- [1] "Descritor interpretive qui-quadrado p-valor: 0"
- [1] "Descritor interview qui-quadrado p-valor: 0"

[1] "Descritor interviews qui-quadrado p-valor: 0"

[1] "Descritor investigate qui-quadrado p-valor: 0"

[1] "Descritor investigative qui-quadrado p-valor: 0,69"

[1] "Descritor longitudinal qui-quadrado p-valor: 0"

[1] "Descritor math qui-quadrado p-valor: 0"

[1] "Descritor math test qui-quadrado p-valor: 0,47"

[1] "Descritor meta-analysis qui-quadrado p-valor: 0"

[1] "Descritor modeling qui-quadrado p-valor: 0,00002"

[1] "Descritor normative qui-quadrado p-valor: 0,021"

[1] "Descritor observational qui-quadrado p-valor: 0"

[1] "Descritor operationalization qui-quadrado p-valor: 0"

[1] "Descritor phenomenology qui-quadrado p-valor: 0,0053"

[1] "Descritor pilot qui-quadrado p-valor: 0"

[1] "Descritor pilot study qui-quadrado p-valor: 0"

[1] "Descritor problem-oriented qui-quadrado p-valor: Sem ocorrência de uma das categorias"

[1] "Descritor problem-solving qui-quadrado p-valor: 0,026"

[1] "Descritor process analysis qui-quadrado p-valor: Sem ocorrência de uma das categorias"

[1] "Descritor processual qui-quadrado p-valor: 0,56"

[1] "Descritor proof qui-quadrado p-valor: 0"

[1] "Descritor protocol analysis qui-quadrado p-valor: Sem ocorrência de uma das categorias"

[1] "Descritor prototype qui-quadrado p-valor: 0"

[1] "Descritor pure qui-quadrado p-valor: 0"

[1] "Descritor qualitative qui-quadrado p-valor: 0"

[1] "Descritor quantitative qui-quadrado p-valor: 0"

[1] "Descritor real qui-quadrado p-valor: 0"

[1] "Descritor real-world qui-quadrado p-valor: 0"

[1] "Descritor real world qui-quadrado p-valor: 0,3"

[1] "Descritor rebuttal qui-quadrado p-valor: 0,16"

[1] "Descritor replication qui-quadrado p-valor: 0"

[1] "Descritor scientific qui-quadrado p-valor: 0"

[1] "Descritor secondary data qui-quadrado p-valor: 0,00016"

[1] "Descritor simulated qui-quadrado p-valor: 0"

[1] "Descritor simulation qui-quadrado p-valor: 0"

[1] "Descritor speech analysis qui-quadrado p-valor: 0,32"

[1] "Descritor standards qui-quadrado p-valor: 0"

[1] "Descritor static analysis qui-quadrado p-valor: 0"

[1] "Descritor survey qui-quadrado p-valor: 0"

[1] "Descritor systematic literature review qui-quadrado p-valor: 0"

[1] "Descritor theoretical qui-quadrado p-valor: 0"

[1] "Descritor usability qui-quadrado p-valor: 0"

[1] "Descritor verticals qui-quadrado p-valor: Sem ocorrência de uma das categorias"

[1] "Descritor workload qui-quadrado p-valor: 0"

## Resultados por descritores da ontologia

Os resultados expressos na tabela abaixo representam o número de ocorrências obtidas para cada descritor da ontologia de maneira individual, sem qualquer tipo de associação entre eles. Os números também correspondem ao total de aceitos (ordenação da tabela juntamente com a categoria) e não aceitos na classificação pelo algoritmo k-nn. Os números são puros e os percentuais foram calculados com base no total do descritor (%Desc) e ao total geral de ocorrências (%Tot). Novamente note que alguns percentuais não puderam ser representados matematicamente com o número de casas decimais utilizados nesta análise. Reitera-se que tais resultados não foram descartados por comporem os resultados totais por categoria do descritor.

O destaque da tabela cabe à última coluna - Pred. - Predominância do resultado. Ela expressa a comparação

dos resultados aceitos e não aceitos. Se um é maior que o outro. Se o ‘n’ de aceitos é maior que o de não aceitos na coluna será mostrado o conteúdo ‘S’ caso contrário ‘N’. Analisando esta coluna pode-se ver que os descritores ‘verticals’, ‘controled model’ e ‘doctrinal’ os quais, além da baixa ocorrência, não tiveram nenhum caso de classificação positiva. Indicando sua não utilização dentro da pesquisa em computação para a população estudada.

Ainda analisando a coluna predominância, pode-se perceber que quase a totalidade, daqueles descritores onde predominam os não aceitos na classificação, os percentuais em relação ao total de ocorrências são inferiores a 1% (0,6% para ser exato). Excetuando os descritores ‘modeling’ (4,7%), ‘concept’(2,1%) e ‘effectiveness’ (2,07%). Analisando individualmente estes casos nota-se que as diferenças entre aceitos e não aceitos, são relativamente pequenas, quase inexistentes não representando mais que 1,5% em relação ao total do descritor.

Table 4: Aceite por descritores da ontologia

Categoria	Descritor	n	%	Total Sim	%Desc.	%Tot.	Total Não	%Desc.	%Tot.	Pred.
Ambiente	real	6799	5,46	4718	69	3,79	2081	30,6	1,67	S
Ambiente	real-world	3028	2,43	2338	77	1,88	690	22,8	0,55	S
Ambiente	simulated	1496	1,20	1167	78	0,94	329	22,0	0,26	S
Ambiente	real world	304	0,24	161	53	0,13	143	47,0	0,11	S
Ambiente	in vivo	47	0,04	32	68	0,03	15	31,9	0,01	S
Ambiente	in vitro	19	0,02	8	42	0,01	11	57,9	0,01	N
Dados	qualitative	3153	2,53	2995	95	2,40	158	5,0	0,13	S
Dados	quantitative	1762	1,41	1403	80	1,13	359	20,4	0,29	S
Dados	longitudinal	536	0,43	365	68	0,29	171	31,9	0,14	S
Dados	cross-sectional	75	0,06	59	79	0,05	16	21,3	0,01	S
Dados	cross-section	46	0,04	19	41	0,02	27	58,7	0,02	N
Dados	verticals	4	0,00	0	0	0,00	4	100,0	0,00	N
Estratégia de Validação/Coleta de Dados	survey	6692	5,37	6082	91	4,88	610	9,1	0,49	S
Estratégia de Validação/Coleta de Dados	experiment	3978	3,19	2463	62	1,98	1515	38,1	1,22	S
Estratégia de Validação/Coleta de Dados	workload	2616	2,10	2080	80	1,67	536	20,5	0,43	S
Estratégia de Validação/Coleta de Dados	benchmark	2957	2,37	2006	68	1,61	951	32,2	0,76	S
Estratégia de Validação/Coleta de Dados	interviews	2306	1,85	1876	81	1,51	430	18,6	0,35	S
Estratégia de Validação/Coleta de Dados	interview	2073	1,66	1673	81	1,34	400	19,3	0,32	S
Estratégia de Validação/Coleta de Dados	case study	1395	1,12	1236	89	0,99	159	11,4	0,13	S
Estratégia de Validação/Coleta de Dados	pilot	870	0,70	773	89	0,62	97	11,2	0,08	S
Estratégia de Validação/Coleta de Dados	comparative	703	0,56	565	80	0,45	138	19,6	0,11	S
Estratégia de Validação/Coleta de Dados	static analysis	650	0,52	389	60	0,31	261	40,1	0,21	S
Estratégia de Validação/Coleta de Dados	replication	632	0,51	253	40	0,20	379	60,0	0,30	N
Estratégia de Validação/Coleta de Dados	grounded theory	237	0,19	184	78	0,15	53	22,4	0,04	S
Estratégia de Validação/Coleta de Dados	dynamic analysis	273	0,22	170	62	0,14	103	37,7	0,08	S
Estratégia de Validação/Coleta de Dados	pilot study	179	0,14	145	81	0,12	34	19,0	0,03	S
Estratégia de Validação/Coleta de Dados	meta-analysis	150	0,12	132	88	0,11	18	12,0	0,01	S
Estratégia de Validação/Coleta de Dados	action research	72	0,06	51	71	0,04	21	29,2	0,02	S
Estratégia de Validação/Coleta de Dados	phenomenology	29	0,02	22	76	0,02	7	24,1	0,01	S
Estratégia de Validação/Coleta de Dados	secondary data	18	0,01	17	94	0,01	1	5,6	0,00	S
Estratégia de Validação/Coleta de Dados	systematic literature review	135	0,11	15	11	0,01	120	88,9	0,10	N
Estratégia de Validação/Coleta de Dados	math test	17	0,01	7	41	0,01	10	58,8	0,01	N
Estratégia de Validação/Coleta de Dados	case-control	6	0,00	6	100	0,00	0	0,0	0,00	S
Estratégia de Validação/Coleta de Dados	conceptual analysis	6	0,00	6	100	0,00	0	0,0	0,00	S
Estratégia de Validação/Coleta de Dados	community research	10	0,01	4	40	0,00	6	60,0	0,00	N
Estratégia de Validação/Coleta de Dados	protocol analysis	3	0,00	3	100	0,00	0	0,0	0,00	S
Estratégia de Validação/Coleta de Dados	speech analysis	4	0,00	3	75	0,00	1	25,0	0,00	S
Estratégia de Validação/Coleta de Dados	hermeneutics	2	0,00	1	50	0,00	1	50,0	0,00	S
Estratégia de Validação/Coleta de Dados	instrument application	1	0,00	1	100	0,00	0	0,0	0,00	S
Metodologia	simulation	6111	4,90	4182	68	3,36	1929	31,6	1,55	S
Metodologia	modeling	5681	4,56	2681	47	2,15	3000	52,8	2,41	N
Metodologia	prototype	1452	1,17	1388	96	1,11	64	4,4	0,05	S
Metodologia	historical	1365	1,10	1270	93	1,02	95	7,0	0,08	S
Metodologia	observational	483	0,39	422	87	0,34	61	12,6	0,05	S
Metodologia	control model	30	0,02	25	83	0,02	5	16,7	0,00	S
Metodologia	historic	131	0,11	16	12	0,01	115	87,8	0,09	N
Metodologia	process analysis	4	0,00	4	100	0,00	0	0,0	0,00	S
Metodologia	processual	3	0,00	1	33	0,00	2	66,7	0,00	N
Metodologia	controlled model	1	0,00	0	0	0,00	1	100,0	0,00	N
Método	evaluation	9298	7,46	8409	90	6,75	889	9,6	0,71	S
Método	experimental	4924	3,95	3919	80	3,15	1005	20,4	0,81	S
Método	engineering	5658	4,54	3550	63	2,85	2108	37,3	1,69	S
Método	investigate	2397	1,92	2247	94	1,80	150	6,3	0,12	S



Categoria	Descritor	n	%	Total Sim	%Desc.	%Tot.	Total Não	%Desc.	%Tot.	Pred.
Método	formal	3383	2,72	2022	60	1,62	1361	40,2	1,09	S
Método	theoretical	2502	2,01	1515	61	1,22	987	39,5	0,79	S
Método	causal	1589	1,28	1369	86	1,10	220	13,8	0,18	S
Método	scientific	1916	1,54	1281	67	1,03	635	33,1	0,51	S
Método	conceptual	1171	0,94	1095	94	0,88	76	6,5	0,06	S
Método	concept	2531	2,03	775	31	0,62	1756	69,4	1,41	N
Método	analytical	832	0,67	667	80	0,54	165	19,8	0,13	S
Método	exploratory	715	0,57	536	75	0,43	179	25,0	0,14	S
Método	descriptive	765	0,61	482	63	0,39	283	37,0	0,23	S
Método	normative	432	0,35	240	56	0,19	192	44,4	0,15	S
Método	math	497	0,40	164	33	0,13	333	67,0	0,27	N
Método	problem-solving	234	0,19	134	57	0,11	100	42,7	0,08	S
Método	standards	863	0,69	130	15	0,10	733	84,9	0,59	N
Método	constructive	401	0,32	91	23	0,07	310	77,3	0,25	N
Método	investigative	98	0,08	51	52	0,04	47	48,0	0,04	S
Método	evaluative	62	0,05	37	60	0,03	25	40,3	0,02	S
Método	empiric	14	0,01	11	79	0,01	3	21,4	0,00	S
Método	cause-and-effect	8	0,01	8	100	0,01	0	0,0	0,00	S
Método	problem-oriented	1	0,00	1	100	0,00	0	0,0	0,00	S
Método	doctrinal	1	0,00	0	0	0,00	1	100,0	0,00	N
Natureza	applied	6416	5,15	4807	75	3,86	1609	25,1	1,29	S
Natureza	pure	964	0,77	608	63	0,49	356	36,9	0,29	S
Propósito paradigma	behavioral	2202	1,77	1145	52	0,92	1057	48,0	0,85	S
Propósito paradigma	interpretive	137	0,11	100	73	0,08	37	27,0	0,03	S
Propósito paradigma	critique	400	0,32	77	19	0,06	323	80,8	0,26	N
Propósito paradigma	design science	37	0,03	27	73	0,02	10	27,0	0,01	S
Prova Científica Secundária	proof	6969	5,59	6739	97	5,41	230	3,3	0,18	S
Prova Científica Secundária	rebuttal	8	0,01	6	75	0,00	2	25,0	0,00	S
Prova Científica	induction	1131	0,91	979	87	0,79	152	13,4	0,12	S
Prova Científica	deduction	64	0,05	34	53	0,03	30	46,9	0,02	S
Prova Científica	abduction	39	0,03	31	79	0,02	8	20,5	0,01	S
Tipo de Validação do Resultado	efficiency	3194	2,56	2586	81	2,08	608	19,0	0,49	S
Tipo de Validação do Resultado	usability	1546	1,24	1345	87	1,08	201	13,0	0,16	S
Tipo de Validação do Resultado	effectiveness	2499	2,01	640	26	0,51	1859	74,4	1,49	N
Tipo de Validação do Resultado	cost-effectiveness	86	0,07	79	92	0,06	7	8,1	0,01	S
Tipo de Validação do Resultado	operationalization	73	0,06	65	89	0,05	8	11,0	0,01	S
NA	Total	124601	100,00	91419	NA	73,37	33182	NA	26,63	NA

## Resultados por categoria da ontologia

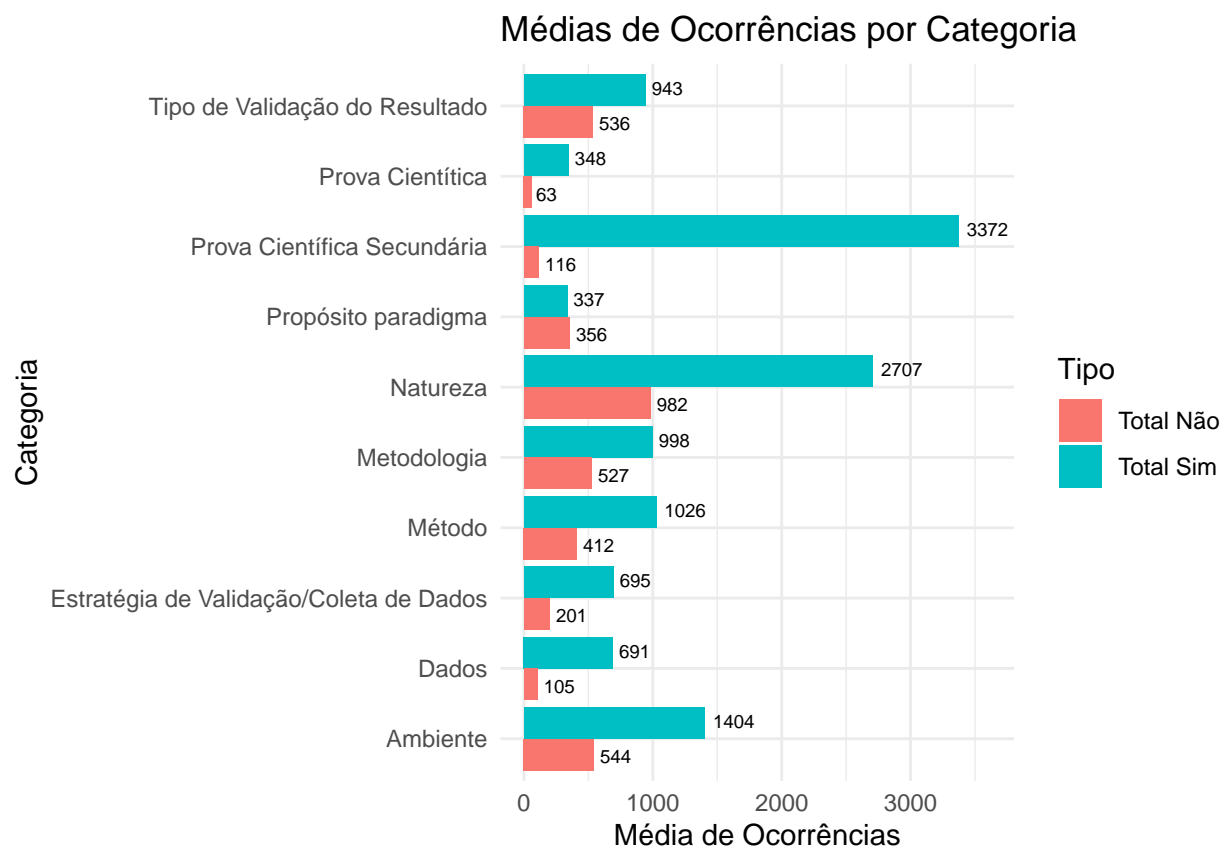
Para os resultados por categoria da ontologia não considerou-se o ‘n’ puro, uma vez que as categorias possuem, internamente, números distintos de descritores. Isto faria com que uma categoria com maior número de descritores, logicamente, tivesse maior número de ocorrências e por consequência valores enfiados em suas contagens. Nesta parte da análise foram considerados os número de descritores por categoria, sendo o número de ocorrências dividido por tais valores componto assim um valor para ‘n’ normalizado ou a média interna por categoria.

Exemplificando, percebe-se que a categoria com maior ocorrências brutas é ‘Método’ com um total de 39.429 ocorrências e 28 descritores. Ao se normalizar este valor obtém-se um ‘n’ normalizado de 1408. Comparando-se com a categoria ‘Natureza’, esta possui apenas 2 descritores resultando em um ‘n’ de 3690. Note que o objetivo desta análise não é estabelecer uma comparação entre as categorias, tal ação é apenas para exemplificar e justificar a forma como os números foram analisados.

Dito isto, os demais dados da tabela foram calculados levando-se em consideração o ‘n’ normalizado ou média por categoria. Nesta tabela não foi calculada a predominância dos valores, uma vez que, exceto a categoria ‘Propósito paradigma’, todas as demais possuem predominância positiva. A categoria em questão possui uma diferença entre os valores para aceitar e não aceitos percentualmente nula inferior a 0,01% ou 19 ocorrências.

Table 5: Aceite por categorias e descritores da ontologia

Categoria	N.Desc.	n	%	n.Norm.	Total Sim	%SCat.	%STot.	Total Não	%NCat.	%NTot.
Ambiente	6	11693	9,38	1949	1404	12,0	1,13	544	4,65	0,44
Dados	7	5576	4,48	797	691	12,4	0,55	105	1,88	0,08
Estratégia de Validação/Coleta de Dados	29	26014	20,88	897	695	2,7	0,56	201	0,77	0,16
Metodologia	10	15261	12,25	1526	998	6,5	0,80	527	3,45	0,42
Método	28	40292	32,34	1439	1026	2,5	0,82	412	1,02	0,33
Natureza	2	7380	5,92	3690	2707	36,7	2,17	982	13,31	0,79
Propósito paradigma	4	2776	2,23	694	337	12,1	0,27	356	12,82	0,29
Prova Científica Secundária	2	6977	5,60	3488	3372	48,3	2,71	116	1,66	0,09
Prova Científica	3	1234	0,99	411	348	28,2	0,28	63	5,11	0,05
Tipo de Validação do Resultado	5	7398	5,94	1480	943	12,8	0,76	536	7,25	0,43
Total	96	124601	100,00	1298	12521	NA	10,05	3842	NA	3,08



## Resultados individuais por categoria

A seguir, as categorias são sumarizadas individualmente e seus destaques apresentados.

### Categoria - Ambiente

Maior número de ocorrências descritor: real , com média 1133 . Com percentual de aceitação 3,8 % e não aceitação de 1,7 %.

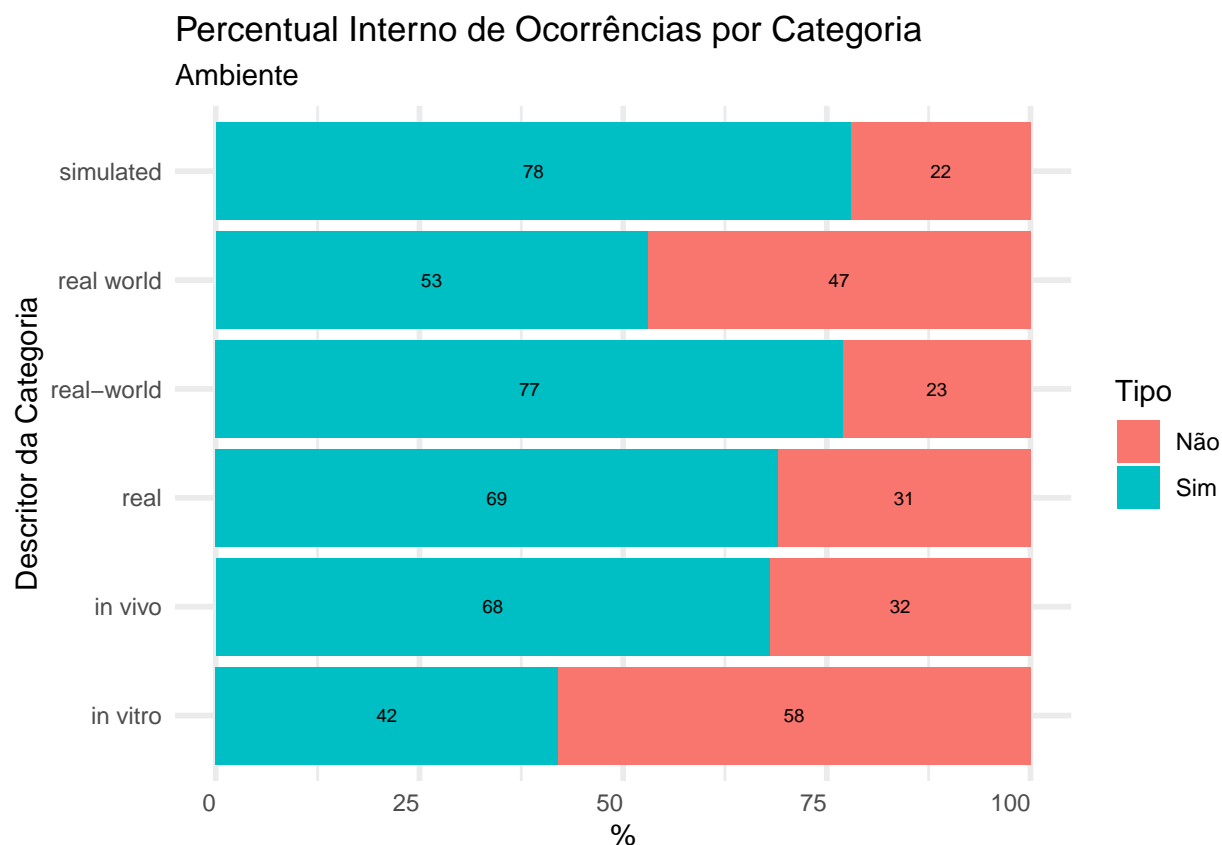
Menor número de ocorrências descritor: in vitro , com média 3,2 . Com percentual de aceitação 0,0064 % e não aceitação de 0,0088 %.

Maior percentual de aceitação sobre o total, descritor: real , com média 1133 . Com percentual de aceitação 3,8 % e não aceitação de 1,7 %.

Menor percentual de aceitação sobre o total, descritor: in vitro , com média 3,2 . Com percentual de aceitação 0,0064 % e não aceitação de 0,0088 %.

Table 6: Distribuição dos resultados individuais da categoria: Ambiente

Descritor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
in vitro	19	0,02	3,2	8	42	0,01	11	58	0,01
in vivo	47	0,04	7,8	32	68	0,03	15	32	0,01
real	6799	5,46	1133,2	4718	69	3,79	2081	31	1,67
real world	304	0,24	50,7	161	53	0,13	143	47	0,11
real-world	3028	2,43	504,7	2338	77	1,88	690	23	0,55
simulated	1496	1,20	249,3	1167	78	0,94	329	22	0,26



### Categoria - Dados

Maior número de ocorrências descritor: qualitative , com média 450 . Com percentual de aceitação 2,4 % e não aceitação de 0,13 %.

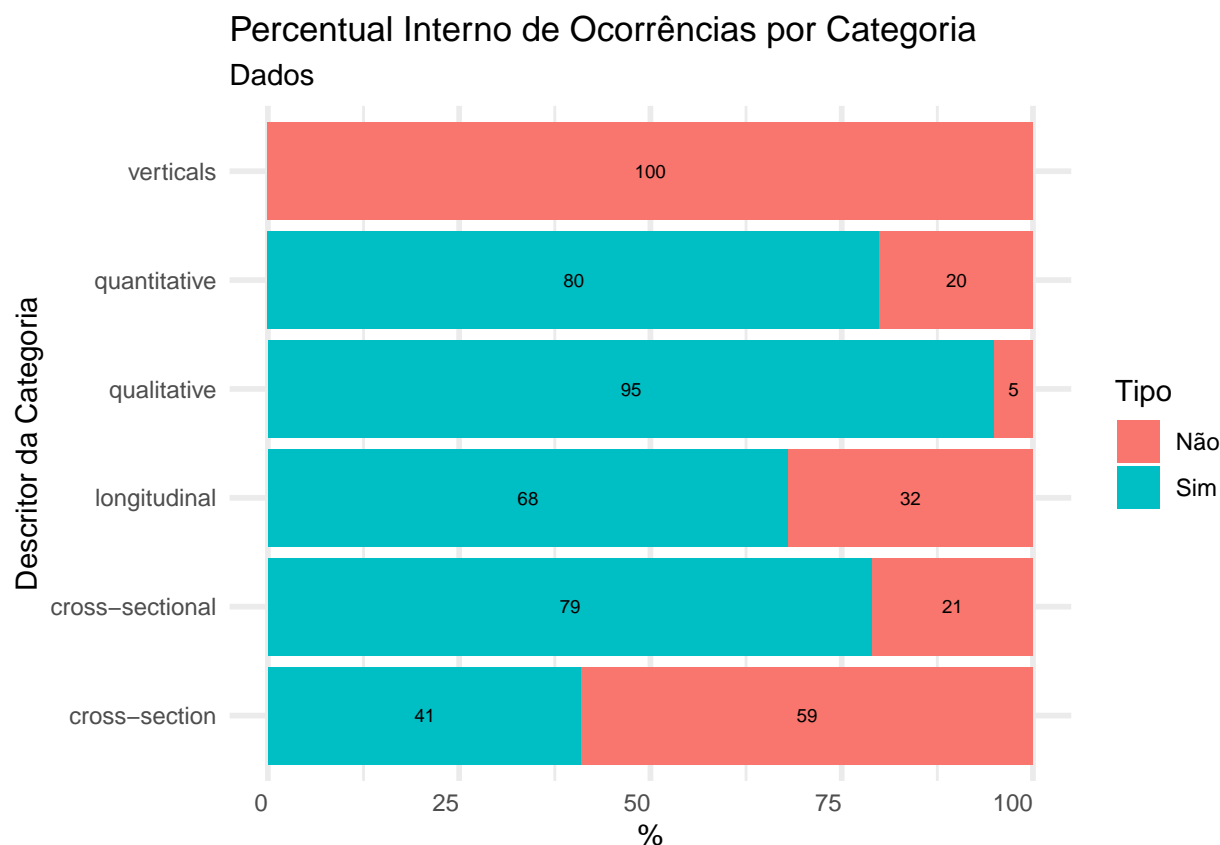
Menor número de ocorrências descritor: verticals , com média 0,57 . Com percentual de aceitação 0 % e não aceitação de 0,0032 %.

Maior percentual de aceitação sobre o total, descritor: qualitative , com média 450 . Com percentual de aceitação 2,4 % e não aceitação de 0,13 %.

Menor percentual de aceitação sobre o total, descritor: verticals , com média 0,57 . Com percentual de aceitação 0 % e não aceitação de 0,0032 %.

Table 7: Distribuição dos resultados individuais da categoria: Dados

Descritor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
cross-section	46	0,04	6,57	19	41	0,02	27	59	0,02
cross-sectional	75	0,06	10,71	59	79	0,05	16	21	0,01
longitudinal	536	0,43	76,57	365	68	0,29	171	32	0,14
qualitative	3153	2,53	450,43	2995	95	2,40	158	5	0,13
quantitative	1762	1,41	251,71	1403	80	1,13	359	20	0,29
verticals	4	0,00	0,57	0	0	0,00	4	100	0,00



#### Categoria - Estratégia de Validação/Coleta de Dados

Maior número de ocorrências descritor: survey , com média 231 . Com percentual de aceitação 4,9 % e não aceitação de 0,49 %.

Menor número de ocorrências descritor: instrument application , com média 0,034 . Com percentual de aceitação 8e-04 % e não aceitação de 0 %.

Maior percentual de aceitação sobre o total, descritor: survey , com média 231 . Com percentual de aceitação 4,9 % e não aceitação de 0,49 %.

Menor percentual de aceitação sobre o total, descritor: case-control , com média 0,21 . Com percentual de aceitação 0,0048 % e não aceitação de 0 %.

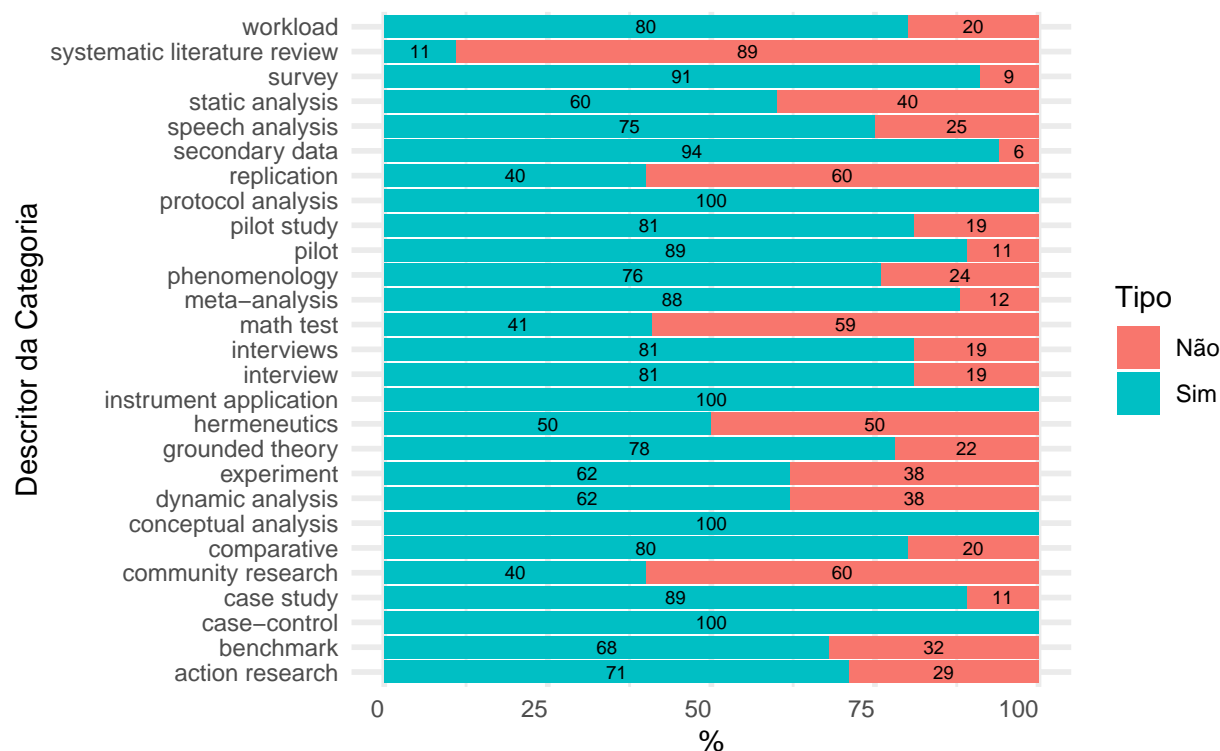
Table 8: Distribuição dos resultados individuais da categoria: Estratégia de Validação/Coleta de Dados

Descritor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
action research	72	0,06	2,48	51	71	0,04	21	29	0,02
benchmark	2957	2,37	101,97	2006	68	1,61	951	32	0,76
case study	1395	1,12	48,10	1236	89	0,99	159	11	0,13
case-control	6	0,00	0,21	6	100	0,00	0	0	0,00
community research	10	0,01	0,34	4	40	0,00	6	60	0,00
comparative	703	0,56	24,24	565	80	0,45	138	20	0,11
conceptual analysis	6	0,00	0,21	6	100	0,00	0	0	0,00
dynamic analysis	273	0,22	9,41	170	62	0,14	103	38	0,08
experiment	3978	3,19	137,17	2463	62	1,98	1515	38	1,22
grounded theory	237	0,19	8,17	184	78	0,15	53	22	0,04
hermeneutics	2	0,00	0,07	1	50	0,00	1	50	0,00
instrument application	1	0,00	0,03	1	100	0,00	0	0	0,00
interview	2073	1,66	71,48	1673	81	1,34	400	19	0,32
interviews	2306	1,85	79,52	1876	81	1,51	430	19	0,35
math test	17	0,01	0,59	7	41	0,01	10	59	0,01
meta-analysis	150	0,12	5,17	132	88	0,11	18	12	0,01
phenomenology	29	0,02	1,00	22	76	0,02	7	24	0,01
pilot	870	0,70	30,00	773	89	0,62	97	11	0,08
pilot study	179	0,14	6,17	145	81	0,12	34	19	0,03
protocol analysis	3	0,00	0,10	3	100	0,00	0	0	0,00
replication	632	0,51	21,79	253	40	0,20	379	60	0,30

Descritor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
secondary data	18	0,01	0,62	17	94	0,01	1	6	0,00
speech analysis	4	0,00	0,14	3	75	0,00	1	25	0,00
static analysis	650	0,52	22,41	389	60	0,31	261	40	0,21
survey	6692	5,37	230,76	6082	91	4,88	610	9	0,49
systematic literature review	135	0,11	4,66	15	11	0,01	120	89	0,10
workload	2616	2,10	90,21	2080	80	1,67	536	20	0,43

## Percentual Interno de Ocorrências por Categoria

### Estratégia de Validação/Coleta de Dados



### Categoria - Metodologia

Maior número de ocorrências descritor: simulation , com média 611 . Com percentual de aceitação 3,4 % e não aceitação de 1,5 %.

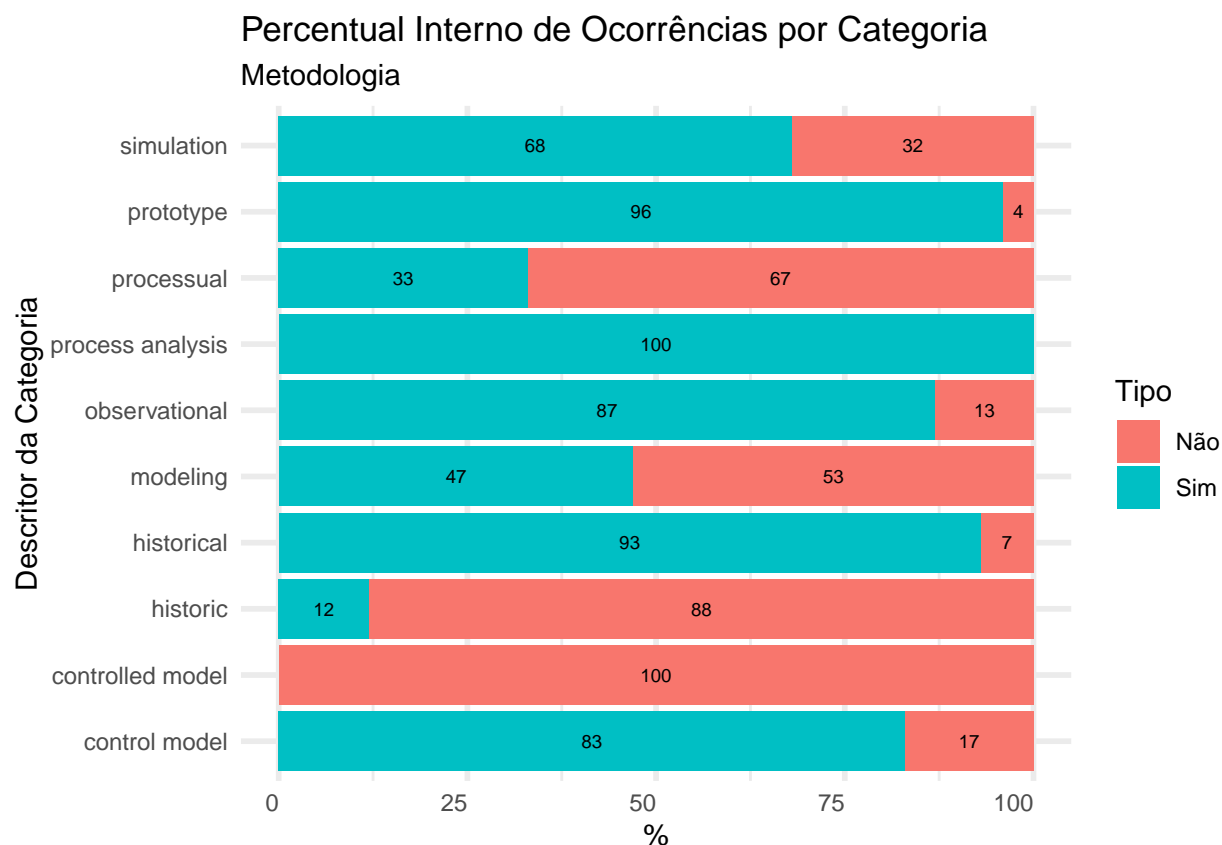
Menor número de ocorrências descritor: controlled model , com média 0,1 . Com percentual de aceitação 0 % e não aceitação de 8e-04 %.

Maior percentual de aceitação sobre o total, descritor: simulation , com média 611 . Com percentual de aceitação 3,4 % e não aceitação de 1,5 %.

Menor percentual de aceitação sobre o total, descritor: process analysis , com média 0,4 . Com percentual de aceitação 0,0032 % e não aceitação de 0 %.

Table 9: Distribuição dos resultados individuais da categoria: Metodologia

Descritor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
control model	30	0,02	3,0	25	83	0,02	5	17	0,00
controlled model	1	0,00	0,1	0	0	0,00	1	100	0,00
historic	131	0,11	13,1	16	12	0,01	115	88	0,09
historical	1365	1,10	136,5	1270	93	1,02	95	7	0,08
modeling	5681	4,56	568,1	2681	47	2,15	3000	53	2,41
observational	483	0,39	48,3	422	87	0,34	61	13	0,05
process analysis	4	0,00	0,4	4	100	0,00	0	0	0,00
processual	3	0,00	0,3	1	33	0,00	2	67	0,00
prototype	1452	1,17	145,2	1388	96	1,11	64	4	0,05
simulation	6111	4,90	611,1	4182	68	3,36	1929	32	1,55



### Categoria - Método

Maior número de ocorrências descritor: evaluation , com média 332 . Com percentual de aceitação 6,7 % e não aceitação de 0,71 %.

Menor número de ocorrências descritor: doctrinal , com média 0,036 . Com percentual de aceitação 0 % e não aceitação de 8e-04 %.

Maior percentual de aceitação sobre o total, descritor: evaluation , com média 332 . Com percentual de aceitação 6,7 % e não aceitação de 0,71 %.

Menor percentual de aceitação sobre o total, descritor: cause-and-effect , com média 0,29 . Com percentual de aceitação 0,0064 % e não aceitação de 0 %.

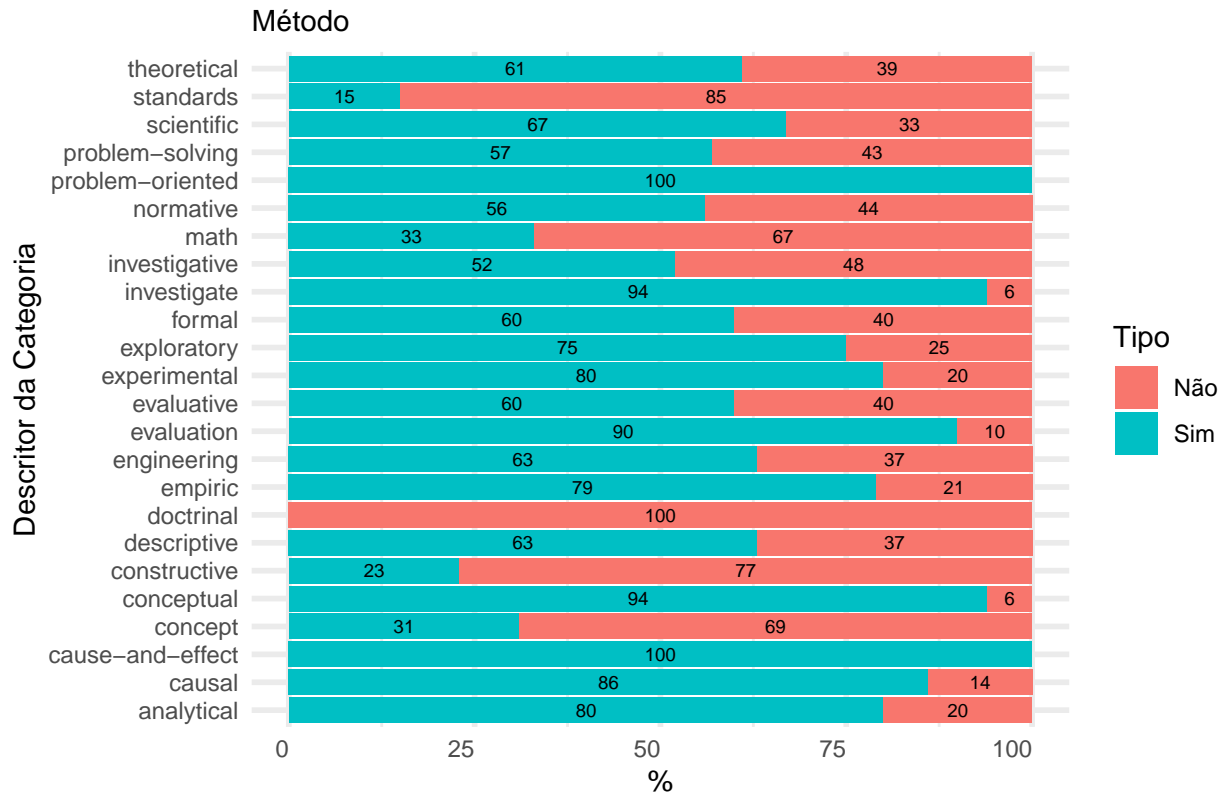
Table 10: Distribuição dos resultados individuais da categoria: Método

Descritor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
analytical	832	0,67	29,71	667	80	0,54	165	20	0,13
causal	1589	1,28	56,75	1369	86	1,10	220	14	0,18
cause-and-effect	8	0,01	0,29	8	100	0,01	0	0	0,00
concept	2531	2,03	90,39	775	31	0,62	1756	69	1,41
conceptual	1171	0,94	41,82	1095	94	0,88	76	6	0,06
constructive	401	0,32	14,32	91	23	0,07	310	77	0,25
descriptive	765	0,61	27,32	482	63	0,39	283	37	0,23
doctrinal	1	0,00	0,04	0	0	0,00	1	100	0,00
empiric	14	0,01	0,50	11	79	0,01	3	21	0,00
engineering	5658	4,54	202,07	3550	63	2,85	2108	37	1,69
evaluation	9298	7,46	332,07	8409	90	6,75	889	10	0,71
evaluative	62	0,05	2,21	37	60	0,03	25	40	0,02
experimental	4924	3,95	175,86	3919	80	3,15	1005	20	0,81
exploratory	715	0,57	25,54	536	75	0,43	179	25	0,14
formal	3383	2,72	120,82	2022	60	1,62	1361	40	1,09
investigate	2397	1,92	85,61	2247	94	1,80	150	6	0,12
investigative	98	0,08	3,50	51	52	0,04	47	48	0,04
math	497	0,40	17,75	164	33	0,13	333	67	0,27
normative	432	0,35	15,43	240	56	0,19	192	44	0,15
problem-oriented	1	0,00	0,04	1	100	0,00	0	0	0,00
problem-solving	234	0,19	8,36	134	57	0,11	100	43	0,08
scientific	1916	1,54	68,43	1281	67	1,03	635	33	0,51



Descritor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
standards	863	0,69	30,82	130	15	0,10	733	85	0,59
theoretical	2502	2,01	89,36	1515	61	1,22	987	39	0,79

## Percentual Interno de Ocorrências por Categoria



### Categoria - Natureza

Maior número de ocorrências descritor: applied , com média 3208 . Com percentual de aceitação 3,9 % e não aceitação de 1,3 %.

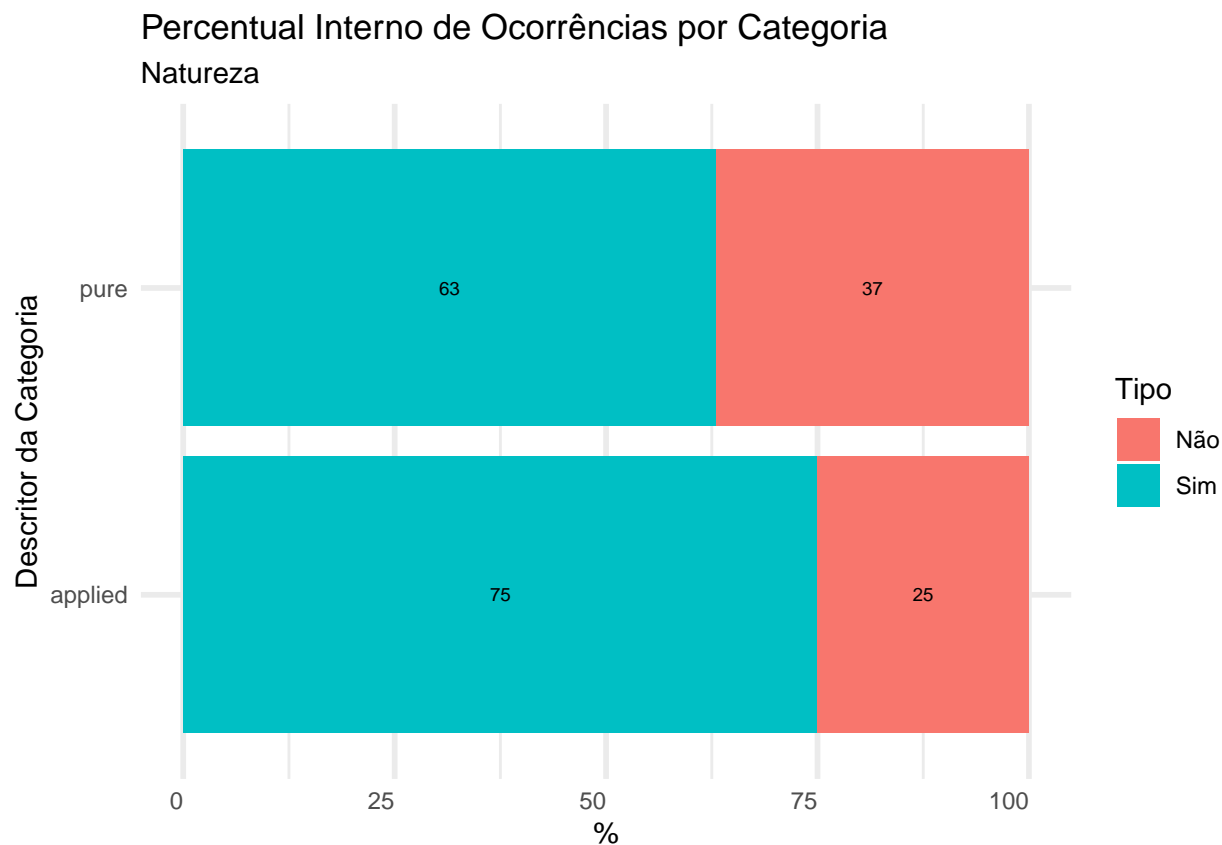
Menor número de ocorrências descritor: pure , com média 482 . Com percentual de aceitação 0,49 % e não aceitação de 0,29 %.

Maior percentual de aceitação sobre o total, descritor: applied , com média 3208 . Com percentual de aceitação 3,9 % e não aceitação de 1,3 %.

Menor percentual de aceitação sobre o total, descritor: pure , com média 482 . Com percentual de aceitação 0,49 % e não aceitação de 0,29 %.

Table 11: Distribuição dos resultados individuais da categoria: Natureza

Descritor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
applied	6416	5,15	3208	4807	75	3,86	1609	25	1,29
pure	964	0,77	482	608	63	0,49	356	37	0,29



### Categoria - Propósito paradigma

Maior número de ocorrências descritor: behavioral , com média 550 . Com percentual de aceitação 0,92 % e não aceitação de 0,85 %.

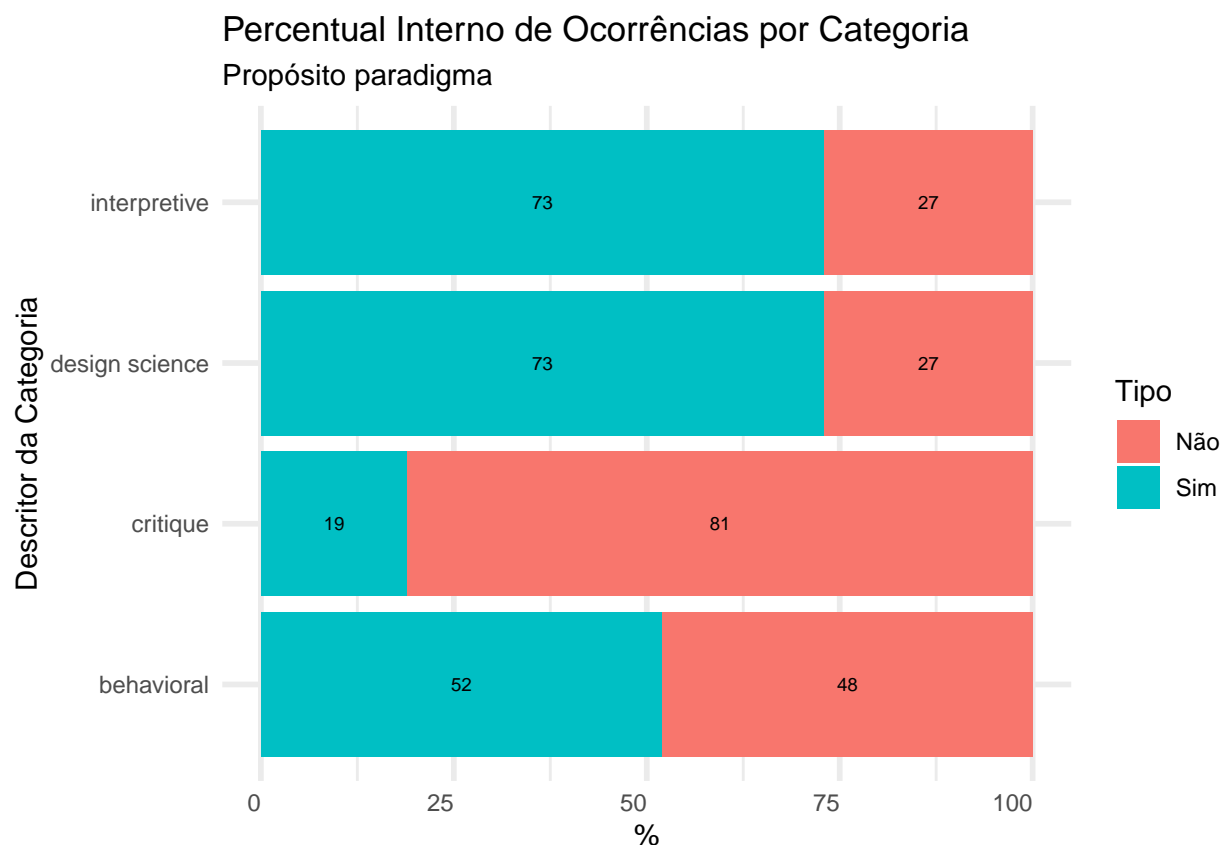
Menor número de ocorrências descritor: design science , com média 9,2 . Com percentual de aceitação 0,022 % e não aceitação de 0,008 %.

Maior percentual de aceitação sobre o total, descritor: behavioral , com média 550 . Com percentual de aceitação 0,92 % e não aceitação de 0,85 %.

Menor percentual de aceitação sobre o total, descritor: design science , com média 9,2 . Com percentual de aceitação 0,022 % e não aceitação de 0,008 %.

Table 12: Distribuição dos resultados individuais da categoria: Propósito paradigma

Descritor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
behavioral	2202	1,77	550,5	1145	52	0,92	1057	48	0,85
critique	400	0,32	100,0	77	19	0,06	323	81	0,26
design science	37	0,03	9,2	27	73	0,02	10	27	0,01
interpretive	137	0,11	34,2	100	73	0,08	37	27	0,03



### Categoria - Prova Científica Secundária

Maior número de ocorrências descritor: proof , com média 3484 . Com percentual de aceitação 5,4 % e não aceitação de 0,18 %.

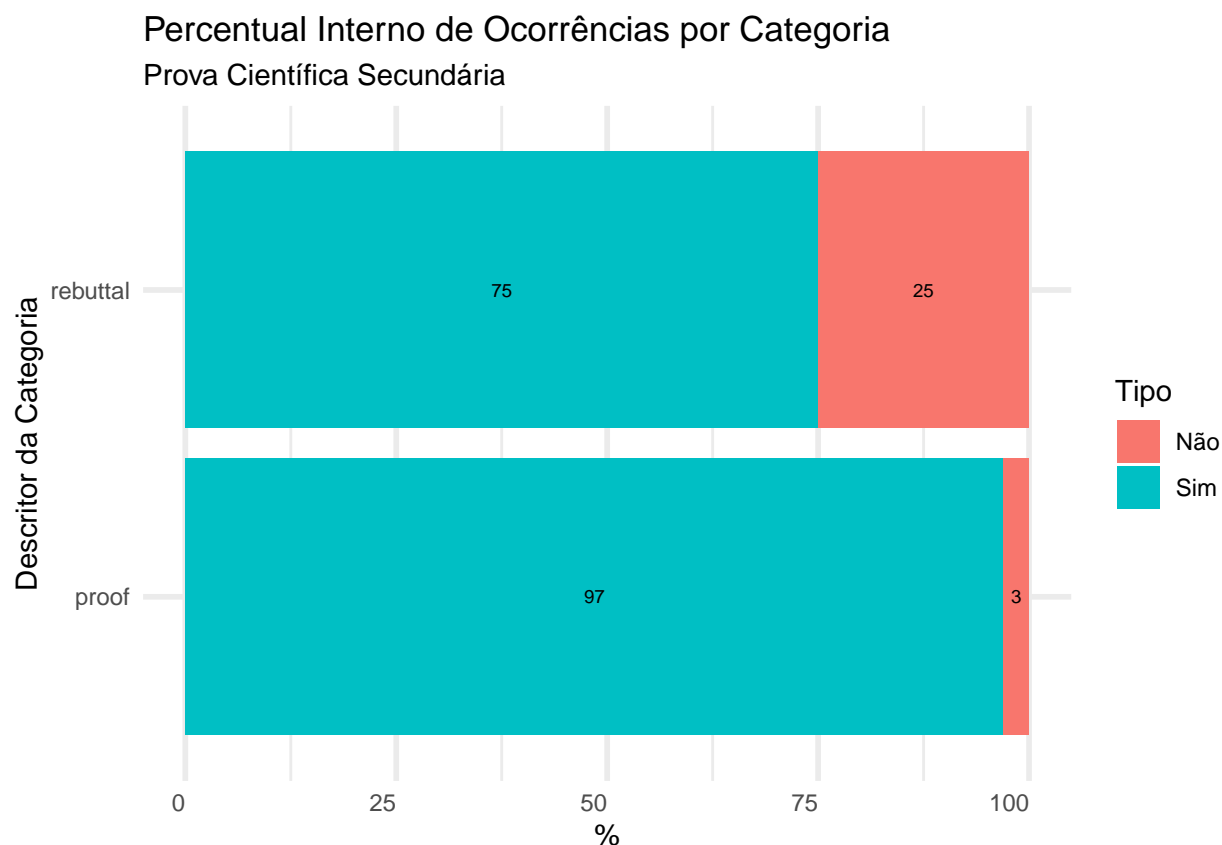
Menor número de ocorrências descritor: rebuttal , com média 4 . Com percentual de aceitação 0,0048 % e não aceitação de 0,0016 %.

Maior percentual de aceitação sobre o total, descritor: proof , com média 3484 . Com percentual de aceitação 5,4 % e não aceitação de 0,18 %.

Menor percentual de aceitação sobre o total, descritor: rebuttal , com média 4 . Com percentual de aceitação 0,0048 % e não aceitação de 0,0016 %.

Table 13: Distribuição dos resultados individuais da categoria: Prova Científica Secundária

Descritor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
proof	6969	5,59	3484	6739	97	5,4	230	3	0,18
rebuttal	8	0,01	4	6	75	0,0	2	25	0,00



### Categoria - Prova Científica

Maior número de ocorrências descritor: induction , com média 377 . Com percentual de aceitação 0,79 % e não aceitação de 0,12 %.

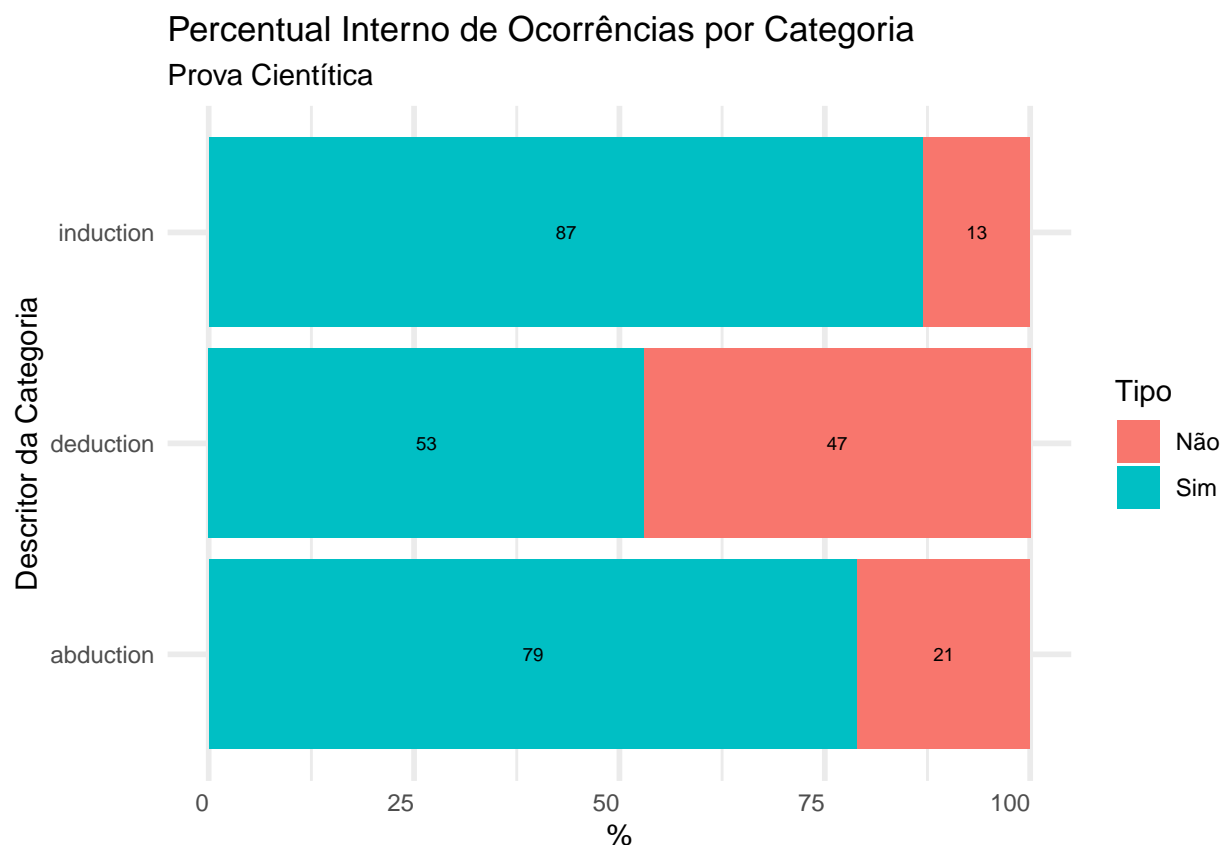
Menor número de ocorrências descritor: abduction , com média 13 . Com percentual de aceitação 0,025 % e não aceitação de 0,0064 %.

Maior percentual de aceitação sobre o total, descritor: induction , com média 377 . Com percentual de aceitação 0,79 % e não aceitação de 0,12 %.

Menor percentual de aceitação sobre o total, descritor: abduction , com média 13 . Com percentual de aceitação 0,025 % e não aceitação de 0,0064 %.

Table 14: Distribuição dos resultados individuais da categoria: Prova Científica

Descritor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
abduction	39	0,03	13	31	79	0,02	8	21	0,01
deduction	64	0,05	21	34	53	0,03	30	47	0,02
induction	1131	0,91	377	979	87	0,79	152	13	0,12



#### Categoria - Tipo de Validação do Resultado

Maior número de ocorrências descritor: efficiency , com média 639 . Com percentual de aceitação 2,1 % e não aceitação de 0,49 %.

Menor número de ocorrências descritor: operationalization , com média 15 . Com percentual de aceitação 0,052 % e não aceitação de 0,0064 %.

Maior percentual de aceitação sobre o total, descritor: efficiency , com média 639 . Com percentual de aceitação 2,1 % e não aceitação de 0,49 %.

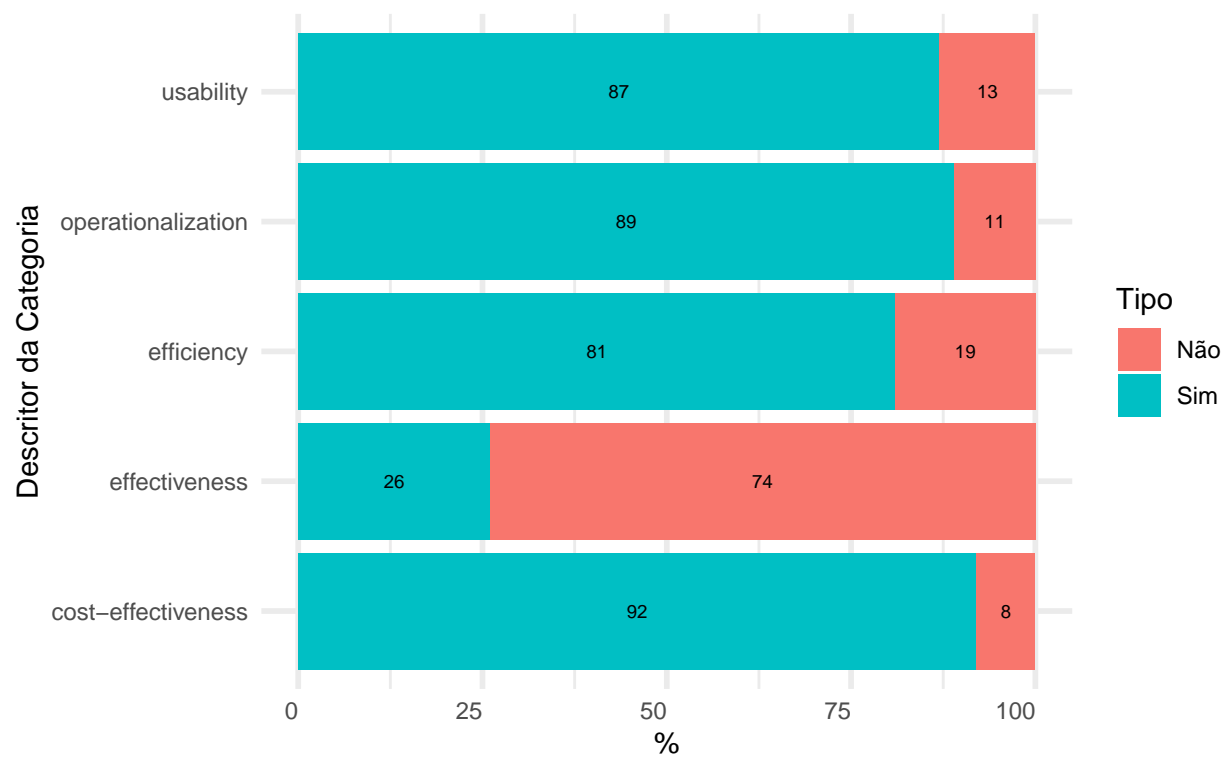
Menor percentual de aceitação sobre o total, descritor: cost-effectiveness , com média 17 . Com percentual de aceitação 0,063 % e não aceitação de 0,0056 %.

Table 15: Distribuição dos resultados individuais da categoria: Tipo de Validação do Resultado

Descritor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
cost-effectiveness	86	0,07	17	79	92	0,06	7	8	0,01
effectiveness	2499	2,01	500	640	26	0,51	1859	74	1,49
efficiency	3194	2,56	639	2586	81	2,08	608	19	0,49
operationalization	73	0,06	15	65	89	0,05	8	11	0,01
usability	1546	1,24	309	1345	87	1,08	201	13	0,16

## Percentual Interno de Ocorrências por Categoria

Tipo de Validação do Resultado



## Tabela sumarizada por Periódico das publicações

No total foram pesquisados 2508 em 68 periódicos da ACM. A tabela abaixo mostra os totais individualizados por periódico ordenados pelo total de ocorrências. Note que a média de ocorrências não esta ligada ao número de artigos. A maior média é do Periódico ACM Trans. Softw. Eng. Methodol. com 146,7 ocorrências em 19 artigos!

Table 16: Distribuição dos resultados por periódico da ACM.

Periódico	Artigos	n	Média	%	Total Sim	%Per.	%Tot.	Total Não	%Per.	%Tot.
Proc. ACM Hum.-Comput. Interact.	522	28263	54,1	22,68	20367	72	16,35	7896	28	6,34
Proc. ACM Program. Lang.	217	12214	56,3	9,80	9552	78	7,67	2662	22	2,14
ACM Trans. Comput.-Hum. Interact.	105	8464	80,6	6,79	5937	70	4,76	2527	30	2,03
ACM Comput. Surv.	65	6387	98,3	5,13	4929	77	3,96	1458	23	1,17
ACM Trans. Graph.	159	5970	37,5	4,79	3890	65	3,12	2080	35	1,67
Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.	113	5447	48,2	4,37	4097	75	3,29	1350	25	1,08
ACM Trans. Model. Comput. Simul.	40	4172	104,3	3,35	2526	61	2,03	1646	39	1,32
ACM Trans. Archit. Code Optim.	79	3254	41,2	2,61	2512	77	2,02	742	23	0,60
ACM/IMS Trans. Data Sci.	73	3253	44,6	2,61	2452	75	1,97	801	25	0,64
Proc. ACM Meas. Anal. Comput. Syst.	57	2856	50,1	2,29	2219	78	1,78	637	22	0,51
ACM Trans. Softw. Eng. Methodol.	19	2821	148,5	2,26	2245	80	1,80	576	20	0,46
ACM Trans. Comput. Educ.	40	2339	58,5	1,88	1680	72	1,35	659	28	0,53
ACM Trans. Program. Lang. Syst.	26	2223	85,5	1,78	1695	76	1,36	528	24	0,42
ACM Trans. Knowl. Discov. Data	41	2143	52,3	1,72	1671	78	1,34	472	22	0,38
J. Hum.-Robot Interact.	31	2099	67,7	1,68	1559	74	1,25	540	26	0,43
J. ACM	33	1941	58,8	1,56	1666	86	1,34	275	14	0,22
SIGPLAN Not.	53	1611	30,4	1,29	1127	70	0,90	484	30	0,39
ACM Trans. Cyber-Phys. Syst.	35	1579	45,1	1,27	1099	70	0,88	480	30	0,39
ACM Trans. Embed. Comput. Syst.	41	1402	34,2	1,13	1020	73	0,82	382	27	0,31
Commun. ACM	72	1377	19,1	1,11	897	65	0,72	480	35	0,39
ACM Trans. Intell. Syst. Technol.	29	1267	43,7	1,02	944	75	0,76	323	25	0,26
J. Emerg. Technol. Comput. Syst.	38	1252	33,0	1,00	906	72	0,73	346	28	0,28
ACM Trans. Interact. Intell. Syst.	19	1248	65,7	1,00	935	75	0,75	313	25	0,25
ACM Trans. Storage	26	1177	45,3	0,94	852	72	0,68	325	28	0,26
ACM Trans. Des. Autom. Electron. Syst.	32	1107	34,6	0,89	824	74	0,66	283	26	0,23
ACM Trans. Priv. Secur.	21	1036	49,3	0,83	792	76	0,64	244	24	0,20
ACM Trans. Sen. Netw.	22	964	43,8	0,77	708	73	0,57	256	27	0,21
Proc. ACM Comput. Graph. Interact. Tech.	44	946	21,5	0,76	668	71	0,54	278	29	0,22
ACM Trans. Comput. Syst.	21	914	43,5	0,73	718	79	0,58	196	21	0,16
ACM Trans. Appl. Percept.	10	860	86,0	0,69	681	79	0,55	179	21	0,14
ACM Trans. Access. Comput.	15	824	54,9	0,66	626	76	0,50	198	24	0,16
ACM Trans. Math. Softw.	25	736	29,4	0,59	498	68	0,40	238	32	0,19
ACM Trans. Multimedia Comput. Commun. Appl.	19	730	38,4	0,59	565	77	0,45	165	23	0,13
ACM Trans. Manage. Inf. Syst.	17	725	42,6	0,58	522	72	0,42	203	28	0,16
ACM Trans. Parallel Comput.	15	697	46,5	0,56	511	73	0,41	186	27	0,15
ACM Trans. Spatial Algorithms Syst.	16	681	42,6	0,55	526	77	0,42	155	23	0,12
ACM Trans. Database Syst.	10	638	63,8	0,51	516	81	0,41	122	19	0,10
ACM Trans. Internet Things	8	541	67,6	0,43	366	68	0,29	175	32	0,14
ACM Trans. Model. Perform. Eval. Comput. Syst.	10	513	51,3	0,41	384	75	0,31	129	25	0,10
Digital Threats	8	507	63,4	0,41	304	60	0,24	203	40	0,16
ACM Trans. Algorithms	24	499	20,8	0,40	421	84	0,34	78	16	0,06
ACM Trans. Inf. Syst.	10	499	49,9	0,40	356	71	0,29	143	29	0,11
ACM Trans. Econ. Comput.	17	497	29,2	0,40	433	87	0,35	64	13	0,05
ACM Trans. Comput. Healthcare	8	460	57,5	0,37	324	70	0,26	136	30	0,11
ACM Trans. Comput. Theory	11	460	41,8	0,37	380	83	0,30	80	17	0,06
ACM Trans. Internet Technol.	11	414	37,6	0,33	307	74	0,25	107	26	0,09
Ubiquity	14	339	24,2	0,27	203	60	0,16	136	40	0,11
SIGCOMM Comput. Commun. Rev.	26	317	12,2	0,25	215	68	0,17	102	32	0,08
Form. Asp. Comput.	3	297	99,0	0,24	231	78	0,19	66	22	0,05
Digit. Gov.: Res. Pract.	13	293	22,5	0,24	219	75	0,18	74	25	0,06
ACM J. Exp. Algorithmics	7	289	41,3	0,23	218	75	0,17	71	25	0,06
ACM Trans. Asian Low-Resour. Lang. Inf. Process.	11	288	26,2	0,23	208	72	0,17	80	28	0,06
J. Data and Information Quality	7	272	38,9	0,22	215	79	0,17	57	21	0,05
SIGMETRICS Perform. Eval. Rev.	31	268	8,6	0,22	166	62	0,13	102	38	0,08
ACM Trans. Evol. Learn. Optim.	4	265	66,2	0,21	171	65	0,14	94	35	0,08
Digital Threats: Research and Practice	5	259	51,8	0,21	169	65	0,14	90	35	0,07
Trans. Soc. Comput.	5	254	50,8	0,20	171	67	0,14	83	33	0,07
ACM Trans. Reconfigurable Technol. Syst.	9	247	27,4	0,20	196	79	0,16	51	21	0,04
J. Comput. Cult. Herit.	4	246	61,5	0,20	190	77	0,15	56	23	0,04
ACM Inroads	10	227	22,7	0,18	143	63	0,11	84	37	0,07
ACM Trans. Web	5	148	29,6	0,12	91	61	0,07	57	39	0,05
ACM Trans. Comput. Logic	3	133	44,3	0,11	118	89	0,09	15	11	0,01
ACM Trans. Auton. Adapt. Syst.	2	107	53,5	0,09	88	82	0,07	19	18	0,02
Interactions	15	103	6,9	0,08	53	51	0,04	50	49	0,04
XRDS	20	97	4,8	0,08	50	52	0,04	47	48	0,04
ACM Transactions on Quantum Computing	2	84	42,0	0,07	54	64	0,04	30	36	0,02
Queue	4	53	13,2	0,04	36	68	0,03	17	32	0,01
SIGACCESS Access. Comput.	1	8	8,0	0,01	7	88	0,01	1	12	0,00
Total	2508	124601	NA	100,00	91419	NA	73,37	33182	NA	26,63

## Tabela Sumarizada por area dos periódicos segundo classificação da SCOPUS para a Ciência

Cada observação na base de dados contém informações a respeito do artigo a que se refere. Dentre tais informações está a variável área, a qual contém as áreas as quais fazem parte do escopo de publicação do periódico. O conteúdo se refere ao periódico e não ao artigo conforme mostra a tabela abaixo.



Nesta tabela é possível ver que mais de uma área estão agrupadas por linha, e, uma mesma área está presente em mais de uma linha. Isto se deve ao fato de, como já descrito, a variável areas se referir ao periódico e não ao artigo.

Tal informação é categorizada segundo a classificação ASJC (All Science Journal Classification Codes - <https://scientificresearch.in/asjc-all-science-journal-classification-codes/>) que é organizada e mantida pela SCOPUS/Elsevier (<https://www.scopus.com/home.uri>) e compreende uma lista das grandes áreas da ciência (da: vida, física, saúde, social e humanas) divididas em 333 sub áreas não consideradas nesta análise.

Table 17: Distribuição segundo o agrupamento de áreas dos periódicos

Areas	n	%	Total Sim	%Cat.	%Tot.	Total Não	%Cat.	%Tot.
Social Sciences (miscellaneous), Human-Computer Interaction, Computer Networks and Communications,	28263	22,68	20367	72	16,35	7896	28	6,34
Software, Safety, Risk, Reliability and Quality,	12214	9,80	9552	78	7,67	2662	22	2,14
Human-Computer Interaction,	8567	6,88	5990	70	4,81	2577	30	2,07
Theoretical Computer Science, General Computer Science,	6387	5,13	4929	77	3,96	1458	23	1,17
Computer Graphics and Computer-Aided Design,	5970	4,79	3890	65	3,12	2080	35	1,67
Human-Computer Interaction, Hardware and Architecture, Computer Networks and Communications,	5447	4,37	4097	75	3,29	1350	25	1,08
Software,	5044	4,05	3940	78	3,16	1104	22	0,89
General Computer Science,	5022	4,03	3726	74	2,99	1296	26	1,04
Modeling and Simulation, Computer Science Applications,	4810	3,86	3461	72	2,78	1349	28	1,08
Human-Computer Interaction, Artificial Intelligence,	4172	3,35	2526	61	2,03	1646	39	1,32
Software, Information Systems, Hardware and Architecture,	3347	2,69	2494	75	2,00	853	25	0,68
Computer Science (miscellaneous), Safety, Risk, Reliability and Quality, Hardware and Architecture, Computer Networks and Communications,	3254	2,61	2512	77	2,02	742	23	0,60
General Computer Science, Education,	2856	2,29	2219	78	1,78	637	22	0,51
Software, Control and Systems Engineering, Information Systems, Hardware and Architecture,	2566	2,06	1823	71	1,46	743	29	0,60
Artificial Intelligence,	1941	1,56	1666	86	1,34	275	14	0,22
Software, Computer Graphics and Computer-Aided Design,	1611	1,29	1127	70	0,90	484	30	0,39
Human-Computer Interaction, Hardware and Architecture, Computer Networks and Communications, Control and Optimization, Artificial Intelligence,	1579	1,27	1099	70	0,88	480	30	0,39
Computer Networks and Communications,	1526	1,22	1106	72	0,89	420	28	0,34
Software, Hardware and Architecture,	1402	1,13	1020	73	0,82	382	27	0,31
Theoretical Computer Science, Artificial Intelligence,	1267	1,02	944	75	0,76	323	25	0,26
Software, Hardware and Architecture, Electrical and Electronic Engineering,	1252	1,00	906	72	0,73	346	28	0,28
Hardware and Architecture,	1177	0,94	852	72	0,68	325	28	0,26
Computer Science Applications, Computer Graphics and Computer-Aided Design, Electrical and Electronic Engineering,	1107	0,89	824	74	0,66	283	26	0,23
General Computer Science, Safety, Risk, Reliability and Quality,	1036	0,83	792	76	0,64	244	24	0,20
Computer Science Applications, Computer Graphics and Computer-Aided Design,	946	0,76	668	71	0,54	278	29	0,22
Theoretical Computer Science, General Computer Science, Experimental and Cognitive Psychology,	860	0,69	681	79	0,55	179	21	0,14
Human-Computer Interaction, Computer Science Applications,	824	0,66	626	76	0,50	198	24	0,16
Software, Applied Mathematics,	736	0,59	498	68	0,40	238	32	0,19
Hardware and Architecture, Computer Networks and Communications,	730	0,59	565	77	0,45	165	23	0,13
Water Science and Technology, Ocean Engineering, Geology,	726	0,58	445	61	0,36	281	39	0,23
Management Information Systems, General Computer Science,	725	0,58	522	72	0,42	203	28	0,16
Software, Modeling and Simulation, Hardware and Architecture, Computer Science Applications, Computational Theory and Mathematics,	697	0,56	511	73	0,41	186	27	0,15
Signal Processing, Information Systems, Modeling and Simulation, Computer Science Applications, Geometry and Topology, Discrete Mathematics and Combinatorics,	681	0,55	526	77	0,42	155	23	0,12
Information Systems,	638	0,51	516	81	0,41	122	19	0,10
Computer Science (miscellaneous), Software, Information Systems, Media Technology, Safety, Risk, Reliability and Quality, Hardware and Architecture, Computer Networks and Communications,	513	0,41	384	75	0,31	129	25	0,10
Information Systems, General Business, Management and Accounting, Computer Science Applications,	499	0,40	356	71	0,29	143	29	0,11
Mathematics (miscellaneous),	499	0,40	421	84	0,34	78	16	0,06
Computer Science (miscellaneous), Statistics and Probability, Economics and Econometrics,	497	0,40	433	87	0,35	64	13	0,05
Marketing, Computational Mathematics,	460	0,37	380	83	0,30	80	17	0,06
Theoretical Computer Science, Computational Theory and Mathematics,	317	0,25	215	68	0,17	102	32	0,08
Software, Computer Networks and Communications,	302	0,24	210	70	0,17	92	30	0,07
Software, Medicine (miscellaneous), Information Systems, Biomedical Engineering, Computer Science Applications, Health Informatics, Health Information Management,	297	0,24	231	78	0,19	66	22	0,05
Software, Theoretical Computer Science,	293	0,24	219	75	0,18	74	25	0,06
Microbiology,	289	0,23	218	75	0,17	71	25	0,06
Theoretical Computer Science,	272	0,22	215	79	0,17	57	21	0,05
Information Systems, Information Systems and Management,	268	0,22	166	62	0,13	102	38	0,08
Software, Hardware and Architecture, Computer Networks and Communications,	246	0,20	190	77	0,15	56	23	0,04
Conservation, Information Systems, Computer Science Applications, Computer Graphics and Computer-Aided Design,	133	0,11	118	89	0,09	15	11	0,01
Theoretical Computer Science, General Computer Science, Logic, Computational Mathematics,	121	0,10	80	66	0,06	41	34	0,03
General Environmental Science, General Earth and Planetary Sciences,	107	0,09	88	82	0,07	19	18	0,02
Control and Systems Engineering, Computer Science (miscellaneous), Software,	40	0,03	28	70	0,02	12	30	0,01
Software, Information Systems, Safety Research, Hardware and Architecture, Computer Science Applications, Computer Networks and Communications,	34	0,03	23	68	0,02	11	32	0,01
General Medicine,	34	0,03	24	71	0,02	10	29	0,01
Ocean Engineering,	124601	100,00	91419	NA	73,37	33182	NA	26,63
Total								

Na tabela abaixo, o conteúdo da variável areas foi desmembrado do agrupamento. Cada linha representa a contagem de artigos em cada area respectivamente. Note que o total não é compatível com o número de artigos pesquisados, isto se deve de que os periódicos são categorizados em mais de uma área e há uma sobreposição na contagem.

Table 18: Distribuição dos resultados por áreas segundo SCOPUS.

Areas	n
Applied Mathematics	25
Artificial Intelligence	147
Biomedical Engineering	4
Computational Mathematics	20
Computational Theory and Mathematics	26
Computer Graphics and Computer-Aided Design	292
Computer Networks and Communications	852
Computer Science (miscellaneous)	86
Computer Science Applications	181
Conservation	4
Control and Optimization	35
Control and Systems Engineering	35
Discrete Mathematics and Combinatorics	16
Economics and Econometrics	17
Education	50
Electrical and Electronic Engineering	70
Experimental and Cognitive Psychology	10
General Business	10
General Computer Science	324
General Earth and Planetary Sciences	2
General Environmental Science	2
General Medicine	1
Geology	12
Geometry and Topology	16
Hardware and Architecture	498
Health Informatics	4
Health Information Management	4
Human-Computer Interaction	855
Information Systems	174
Information Systems and Management	7
Logic	3
Management Information Systems	17
Management and Accounting	10
Marketing	17
Mathematics (miscellaneous)	24
Media Technology	10
Medicine (miscellaneous)	4
Microbiology	13
Modeling and Simulation	71
Ocean Engineering	14
Reliability and Quality	305
Risk	305
Safety	305
Safety Research	1
Signal Processing	16
Social Sciences (miscellaneous)	522
Software	623
Statistics and Probability	17
Theoretical Computer Science	128
Water Science and Technology	12

Áreas	n
Total	6196

Na tabela abaixo foi realizado um exercício para relacionar as classificações de área ASJC, ACM-CCS e Grandes áreas da computação segundo a Sociedade Brasileira de Computação (SBC). Como pode ser visto há uma equivalência entre elas, o que já era esperado uma vez que as três versam sobre a mesma área da ciência.

Table 19: Associação de equivalência das classificações da Ciência da Computação segundo ACM, SCOPUS e SBC.

Classif. ACM	Classif. SCOPUS	Classif. SBC
General and reference	General Computer Science	GA 4 Aplicações da Computação
Hardware	Hardware and Architecture	GA 1 Sistemas Computacionais
Hardware	Signal Processing	GA 1 Sistemas Computacionais
Hardware	Software	GA 2 Sistemas de Software
Computer systems organization	Software	GA 2 Sistemas de Software
Networks	Computer Networks and Communications	GA 1 Sistemas Computacionais
Software and its engineering	Software	GA 2 Sistemas de Software
Theory of computation	Computational Theory and Mathematics	GA 5 Gestão de Dados e de Informações
Mathematics of computing	Computational Theory and Mathematics	GA 2 Sistemas de Software
Information systems	Information Systems	GA 2 Sistemas de Software
Security and privacy	Computer Networks and Communications	GA 1 Sistemas Computacionais
Human-centered computing	Human-Computer Interaction	GA 3 Técnicas e Tecnologias de Computação
Computing methodologies	Artificial Intelligence	GA 3 Técnicas e Tecnologias de Computação
Computing methodologies	Computer Vision and Pattern Recognition	GA 4 Aplicações da Computação
Computing methodologies	Computer Graphics and Computer-Aided Design	GA 4 Aplicações da Computação
Applied computing	Computer Science Applications	GA 4 Aplicações da Computação
Applied computing	Computer Graphics and Computer-Aided Design	GA 4 Aplicações da Computação
Social and professional topics	Computer Science (miscellaneous)	GA 4 Aplicações da Computação

Com base na tabela de equivalência de áreas (ASJC, ACM-CCS e SBC), foi feita a relação entre tal tabela e as áreas desmembradas obtidas nas observações dos dados do presente trabalho, uma vez que estes estão distribuídos segundo a classificação da SCOPUS. Note que no que tange a computação, todas as áreas foram contempladas nos resultados. Ainda é possível notar que o escopo dos periódicos pesquisados não dizem respeito apenas a computação, mas também a outras áreas do conhecimento. Conforme mostra a tabela abaixo.

Table 20: Contagem de artigos por área e associação de áreas segundo SCOPUS, ACM e SBC

Área (SCOPUS)	n	Classif. ACM	Classif. SBC
Applied Mathematics	25	NA	NA
Artificial Intelligence	147	Computing methodologies	GA 3 Técnicas e Tecnologias de Computação
Biomedical Engineering	4	NA	NA
Computational Mathematics	20	NA	NA
Computational Theory and Mathematics	26	Theory of computation	GA 5 Gestão de Dados e de Informações
Computational Theory and Mathematics	26	Mathematics of computing	GA 2 Sistemas de Software
Computer Graphics and Computer-Aided Design	292	Computing methodologies	GA 4 Aplicações da Computação
Computer Graphics and Computer-Aided Design	292	Applied computing	GA 4 Aplicações da Computação
Computer Networks and Communications	852	Networks	GA 1 Sistemas Computacionais
Computer Networks and Communications	852	Security and privacy	GA 1 Sistemas Computacionais
Computer Science (miscellaneous)	86	Social and professional topics	GA 4 Aplicações da Computação
Computer Science Applications	181	Applied computing	GA 4 Aplicações da Computação
Computer Vision and Pattern Recognition	NA	Computing methodologies	GA 4 Aplicações da Computação
Conservation	4	NA	NA
Control and Optimization	35	NA	NA
Control and Systems Engineering	35	NA	NA
Discrete Mathematics and Combinatorics	16	NA	NA
Economics and Econometrics	17	NA	NA
Education	50	NA	NA
Electrical and Electronic Engineering	70	NA	NA
Experimental and Cognitive Psychology	10	NA	NA
General Business	10	NA	NA
General Computer Science	324	General and reference	GA 4 Aplicações da Computação
General Earth and Planetary Sciences	2	NA	NA
General Environmental Science	2	NA	NA
General Medicine	1	NA	NA
Geology	12	NA	NA
Geometry and Topology	16	NA	NA
Hardware and Architecture	498	Hardware	GA 1 Sistemas Computacionais
Health Informatics	4	NA	NA
Health Information Management	4	NA	NA
Human-Computer Interaction	855	Human-centered computing	GA 3 Técnicas e Tecnologias de Computação
Information Systems	174	Information systems	GA 2 Sistemas de Software
Information Systems and Management	7	NA	NA
Logic	3	NA	NA
Management and Accounting	10	NA	NA
Management Information Systems	17	NA	NA
Marketing	17	NA	NA
Mathematics (miscellaneous)	24	NA	NA
Media Technology	10	NA	NA
Medicine (miscellaneous)	4	NA	NA
Microbiology	13	NA	NA
Modeling and Simulation	71	NA	NA
Ocean Engineering	14	NA	NA

Área (SCOPUS)	n	Classif. ACM	Classif. SBC
Reliability and Quality	305	NA	NA
Risk	305	NA	NA
Safety	305	NA	NA
Safety Research	1	NA	NA
Signal Processing	16	Hardware	GA 1 Sistemas Computacionais
Social Sciences (miscellaneous)	522	NA	NA
Software	623	Software and its engineering	GA 2 Sistemas de Software
Software	623	Hardware	GA 2 Sistemas de Software
Software	623	Computer systems organization	GA 2 Sistemas de Software
Statistics and Probability	17	NA	NA
Theoretical Computer Science	128	NA	NA
Total	6196	NA	NA
Water Science and Technology	12	NA	NA

## A tabela de aprendizagem do K-nn

No total, a tabela de aprendizagem possui 6809 entradas ao final da execução do algoritmo de classificação. O que corresponde a 5,46% do total de 124601 observações. A taxa de classificação média para cada regra da tabela de aprendizagem é de 22 observações com uma relação de aproximadamente 5 aceites para 1 descarte.

Table 21: Distribuição dos valores da Tabela de Aprendizagem do K-nn.

Categoria	n	Sim	Não	Taxa	Média
Ambiente	553	10992	3143	3,5	19,9
Dados	421	7894	709	11,1	18,8
Estratégia de Validação/Coleta de Dados	1504	33381	5641	5,9	22,2
Metodologia	1068	14270	5029	2,8	13,4
Método	2177	43533	11109	3,9	20,0
Natureza	189	13550	1894	7,2	71,7
Propósito paradigma	242	1888	1392	1,4	7,8
Prova Científica Secundária	202	15926	229	69,5	78,8
Prova Científica	77	1307	189	6,9	17,0
Tipo de Validação do Resultado	385	6177	2498	2,5	16,0
Total	6818	148918	31833	4,7	21,8

Fim