



Universidade Federal
do Rio de Janeiro

Escola Politécnica

MBA em Inteligência de Negócios

Estatística aplicada

Professora Kate Revoredo

Ana Carolina Rodrigues Vicente
Maycon Leite Ribeiro

Coleta de dados

Em nossa busca pela base de dados para realizar as tarefas, encontramos uma base do governo sobre o programa Mais educação do governo federal, realizamos o trabalho com a base do ano de 2016 pois era mais consistente. Nota-se que nesta base não há ausência de dados em nenhuma das colunas, o que facilita para efetuar algumas análises estatísticas.

Cada registro na base representa uma escola que recebe investimento do governo pelo programa. A base apresenta diversas colunas indicando a localização e suas características. Os atributos que possuem maior representatividade nas análises estatísticas escolhidas são a quantidade de alunos e o total de investimento recebido pela escola. Com estes atributos pode-se recolher informações baseadas em suas localizações e gerar conclusões com as estatísticas aplicadas.

A fonte de dados pode ser encontrada através do link: <http://dados.gov.br/dataset/mec-pme>

Programa Mais Educação - PME

O Programa Mais Educação foi uma estratégia do Ministério da Educação que tem como objetivo melhorar a aprendizagem em língua portuguesa e matemática no ensino fundamental, por meio da ampliação da jornada escolar de crianças e adolescentes, otimizando o tempo de permanência dos estudantes na escola.

As escolas das redes públicas de ensino estaduais, municipais e do Distrito Federal realizam adesão ao Programa e, de acordo com o projeto educativo em curso, optam por desenvolver atividades em alguns macrocampos de acompanhamento pedagógico; educação ambiental; esporte e lazer; direitos humanos em educação; cultura e artes; cultura digital; promoção da saúde; comunicação e uso de mídias; investigação no campo das ciências da natureza e educação econômica.

Conjunto de Dados: Escolas municipais e estaduais com plano de atendimento aprovado no Programa Mais Educação, quantidade de alunos de cada escola, valores (R\$) totais recebidos por adesão, UF e município.

UF_ESCOLA	MUNICIPIO_ESCOLA	ESFERA_GOV	LOCALIZACAO_ESCOLA	NOME_ESCOLA	ALUNADO	TOTAL_RECEBIDO_ADESAO
AC	ACRELANDIA	Estadual	Rural	ESC SANTA LUCIA III	60	R\$6,060.00
AC	BRASILEIA	Estadual	Urbana	ESC INSTITUTO ODILON PRATAGI	40	R\$2,760.00
AM	ANAMA	Estadual	Urbana	ESCOLA ESTADUAL JESUINA REGIS	120	R\$23,880.00

AM	ANAMA	Estadual	Rural	ESCOLA ESTADUAL MARIA NOGUEIRA MARQUES	101	R\$20,781.00
MG	ABADIA DOS DOURADOS	Estadual	Urbana	EE PEDRO ALVARES CABRAL	37	R\$9,195.00
MG	ABRE CAMPO	Municipal	Rural	EM DR SERTORIO DE AMORIM E SILVA	80	R\$24,240.00
RJ	ANGRA DOS REIS	Municipal	Urbana	CEHI JOAO CAROLINO REMEDIOS	93	R\$12,645.00
RJ	ANGRA DOS REIS	Municipal	Rural	CEHI MONSENHOR PINTO DE CARVALHO	44	R\$17,220.00
SP	ADAMANTINA	Estadual	Urbana	FLEURIDES CAVALLINI MENECHINO PROFA	20	R\$1,380.00
SP	ALTINOPOLIS	Municipal	Rural	CARMEM MIGUEL VICARI EMEF	155	R\$12,095.00

**Está é uma amostra da base principal.*

Atributos

Dado	Nome do campo	Tipo	Descrição
UF	UF_ESCOLA	Alfanumérico	UF das Escolas atendidas pelo programa.
Município	MUNICIPIO_ESCOLA	Alfanumérico	Município das Escolas atendidas pelo programa.
Esfera de Governo da Escola atendida	ESFERA_GOVERNO	Alfanumérico	Esfera de Governo da Escola atendida pelo programa (estadual ou municipal).
Localização da Escola atendida	LOCALIZACAO_ESCOLA	Alfanumérico	Tipo da localização da Escola atendida (rural ou urbana).
Nome da Escola atendida	NOME_ESCOLA	Alfanumérico	Nome da Escola atendida pelo programa.
Quantidade de alunos	ALUNADO	Numérico	Quantidade de alunos atendidos pelo programa da Escola.
Total Recebido	TOTAL_RECEBIDO_ADES AO	Numérico	Valor total recebido pela Escola atendida pelo programa.

Dados encontrados nas tarefas

Na execução das tarefas propostas os dados a seguir foram levantados e serviram de base para aplicar estatísticas sobre nossos atributos de principal representatividade.

Dados da Tarefa 2:

Medidas de tendência central	
Atributo:	Alunado
Média	120,2
Mediana	100
Moda	100

Medidas de Dispersão	
Atributo:	Alunado
Amplitude	1410
Desvio médio	69,68
Desvio padrão	105,72
Variância	11175,96

Medidas de intervalo	
Atributo:	Alunado
Mínimo	20
Máximo	1430

Medidas de posição	
Atributo:	Alunado
Quantis 1	55
Quantis 2	100
Quantis 3	150

Medidas de forma	
Atributo:	Alunado
Assimetria	2,99507
Curtose	17,29522

Dados da Tarefa 3:

Distribuição binominal
Para levantamento de dados um pesquisador vai visitar 10 escolas que recebem recursos do mais educação. Sabe-se que a probabilidade do pesquisador encontrar uma escola que recebe mais de R\$ 10.000,00 de investimento pelo programa é de 77,8672%. Qual a probabilidade dele encontrar 5 escolas com mais de R\$ 10.000,00 investido na sua pesquisa?
R: 4,903028%

Dados	
Atributo:	TOTAL RECEBIDO ADESAO
Quantidade total	38766
Quantidade investimento maior que R\$ 10.000,00	30186
Probabilidade do valor ser maior que R\$ 10.000,00 em uma visita	77,86720%
Quantidade menor ou igual a R\$ 10.000,00	8580

Distribuição normal
Em uma amostra da base que possui uma distribuição simétrica de alunos, com desvio padrão de 105,7164 e média de 120,1971. Ao visitar uma das escolas desta amostra, qual a probabilidade de encontrar uma escola que possua mais de 150 alunos?
R: 38,90047 %

Dados	
Atributo	ALUNADO
Assimetria do atributo	2,99507
Desvio Padrão do atributo	105,71640
Média do atributo	120,2
Mediana do atributo	100

Dados da Tarefa 5:

Intervalo de Confiança	
Atributo:	Alunado
Média	120,2
Confiança esperada	95%
Intervalo de Confiança	119,1447, 121,2494

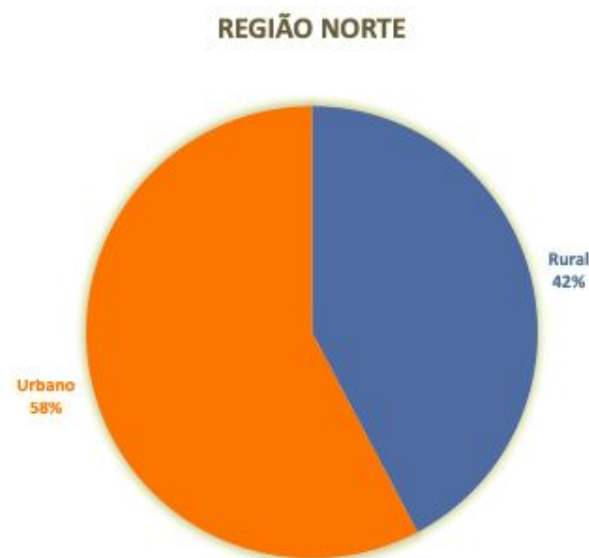
Correlação	
Atributo 1:	Alunado
Atributo 2:	Total Recebido Adesão
Correlação	0,8651064
Inclinação	160,3000000
Intercepto	4.078,600000
Regressão Linear	$Y = 160,3X + 4078,6$

Estatísticas e Gráficos

A base escolhida foi formatada e com base nas informações geradas nas tarefas anteriores foi possível trabalhar os dados da base visando categorizá-los e gerar algumas estatísticas por região, estado e localização da escola. Foi possível também relacionar alguns dados após as análises com o objetivo de tirar conclusões a partir desta relação. Além disso uma correlação forte foi identificada e será apresentada através de um gráfico que vai ilustrar a regressão linear.

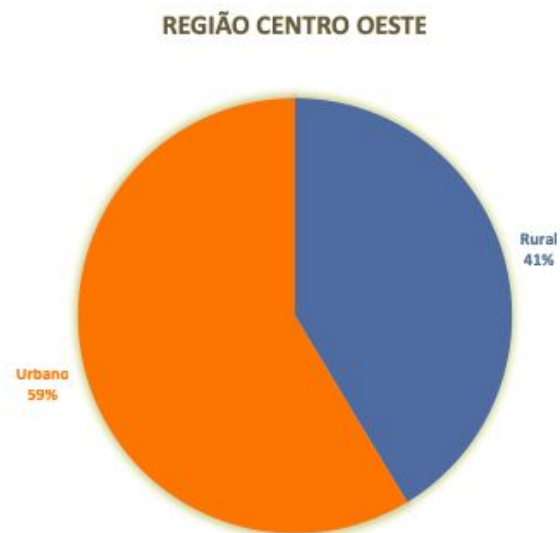
Na região Norte pode-se observar que a média de alunado é maior na localização urbana e também recebe mais investimentos que na localização rural.

Norte	
Média	
Alunado	Total Recebido Adesão
125,877521	R\$24.628,03
Rural	Total Recebido Adesão
89,810736	R\$22.809,82
Urbano	Total Recebido Adesão
122,705882	R\$26.114,67
Total de alunado	
505.650	
Total recebido por regioao	
R\$97.815.888,00	
Grupo	Acre (AC), Amapá (AP), Amazonas (AM), Pará (PA), Rondônia (RO), Roraima (RR) e Tocantins (TO).



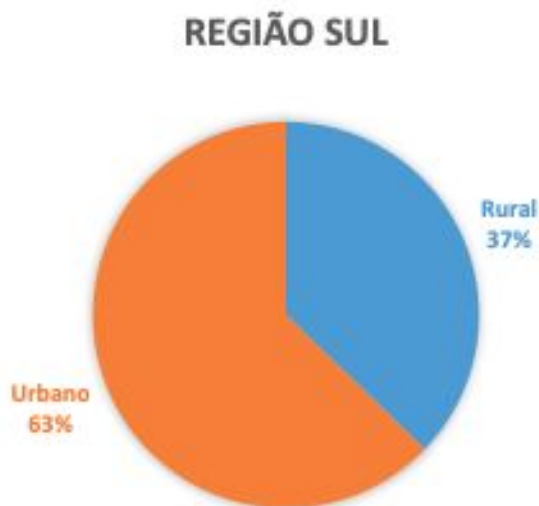
Na região Centro Oeste nota-se um comportamento interessante. A média de alunado da localização urbana é bem maior que na região rural, porém esta diferença não se reflete no total recebido. Pode-se notar que o total recebido entre as localizações é quase equivalente.

Centro Oeste	
Média	
Alunado	Total Recebido Adesão
123,69966	R\$22.932,52
Rural	Total Recebido Adesão
91,41875	R\$22.710,13
Urbano	Total Recebido Adesão
129,63297	R\$22.973,40
Total de alunado	
254.945	
Total recebido por regioao	
R\$46.844.212,00	
Grupo	Goiás (GO), Distrito Federal (DF), Mato Grosso (MT) e Mato Grosso do Sul (MS).



Na região Sul a média de alunado e também o total recebido é maior na localização urbana, porém o recebimento está um pouco próximo também. É provável que as escolas da localização rural receberam um valor por aluno um pouco maior que as escolas urbanas.

Sul	
Média	
Alunado	Total Recebido Adesão
96,3485	R\$17.979,58
Rural	Total Recebido Adesão
63,3228512	R\$16.953,58
Urbano	Total Recebido Adesão
106,692055	R\$18.300,93
Total de alunado	
385.394	
Total recebido por região	
R\$71.219.920,00	
Grupo	Paraná (PR), Santa Catarina (SC) e Rio Grande do Sul (RS)



No Nordeste a média de alunado atendido pelo programa é quase o dobro na localização urbana comparado a localização rural. E mais uma vez pode-se notar que o recebimento não possui esta diferença.

Nordeste	
Média	
Alunado	Total Recebido Adesão
123,007052	R\$26.421,85
Rural	Total Recebido Adesão
84,305627	R\$24.169,30
Urbano	Total Recebido Adesão
160,569273	R\$28.611,39
Total de alunado	
2.232.701	
Total recebido por regioao	
R\$478.396.631,00	
Grupo	Alagoas (AL), Bahia (BA), Ceará (CE), Maranhão (MA), Paraíba (PB), Piauí (PI), Pernambuco (PE), Rio Grande do Norte (RN) e Sergipe (SE).

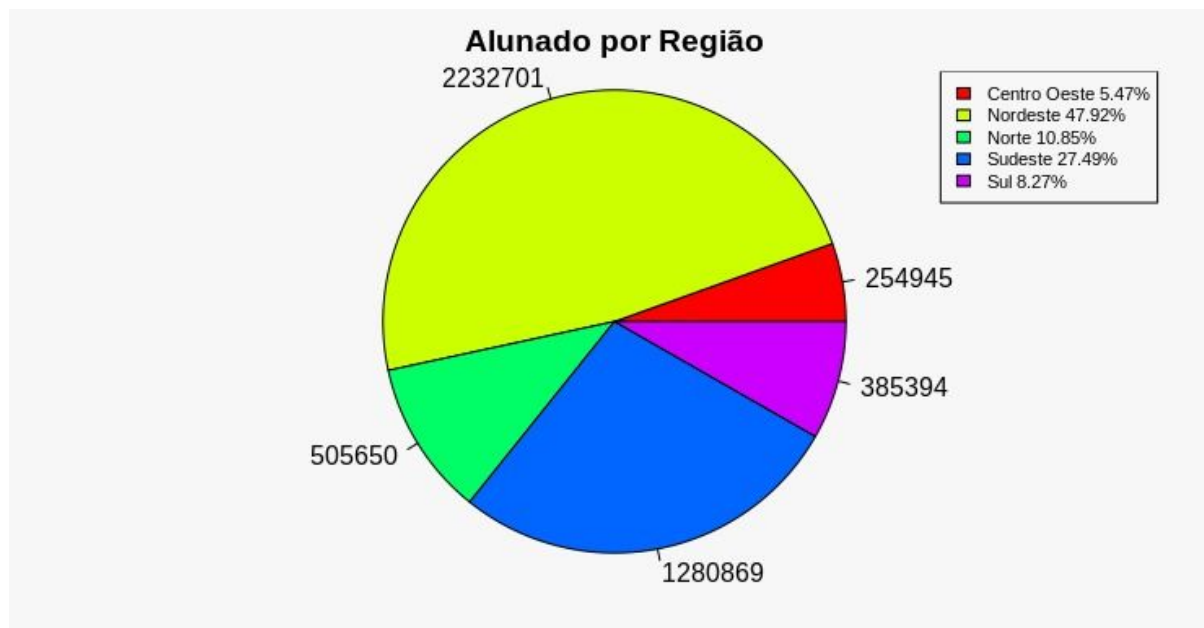


No Sudeste nota-se mais uma vez que a diferença de alunado entre as localizações não reflete na proporção de total recebido.

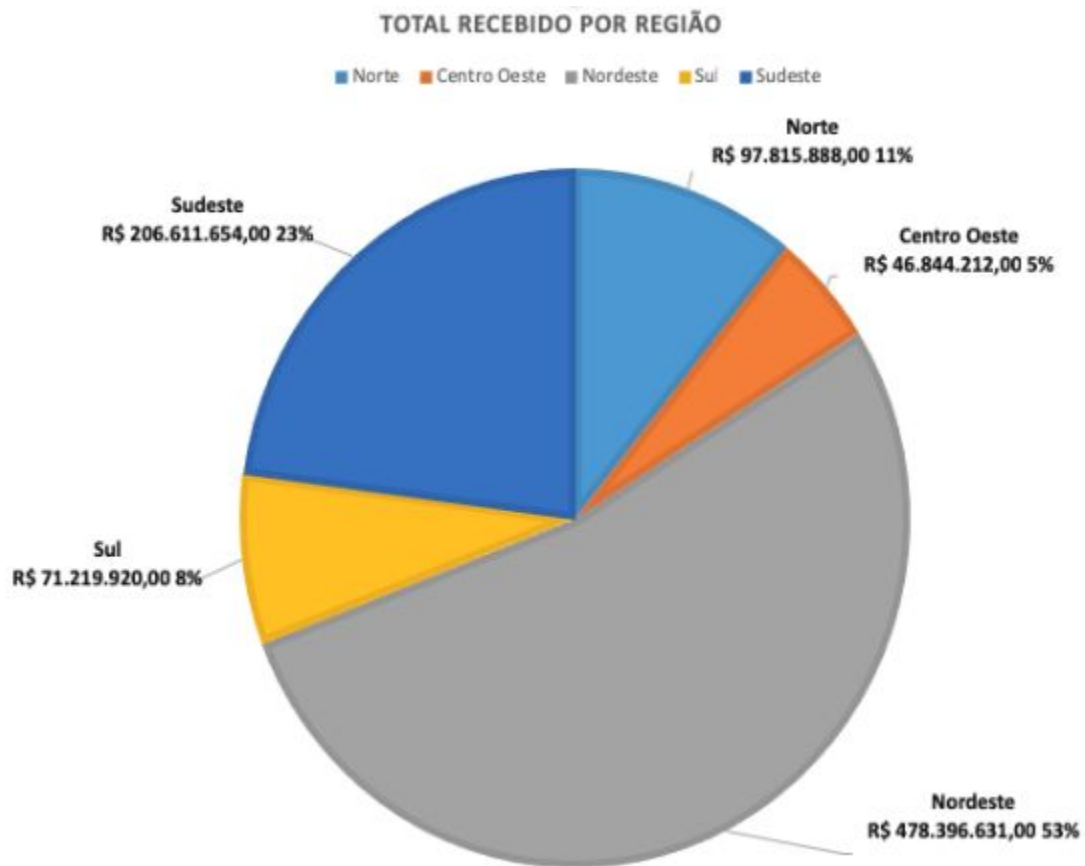
Sudeste	
Média	
Alunado	Total Recebido Adesão
119,927033	R\$20.061,34
Rural	Total Recebido Adesão
73,9306476	R\$19.914,78
Urbano	Total Recebido Adesão
128,846139	R\$20.089,76
Total de alunado	
1.280.869	
Total recebido por regioao	
206.611.654,00	
Grupo	São Paulo (SP), Minas Gerais (MG), Rio de Janeiro (RJ) e Espírito Santo (ES)



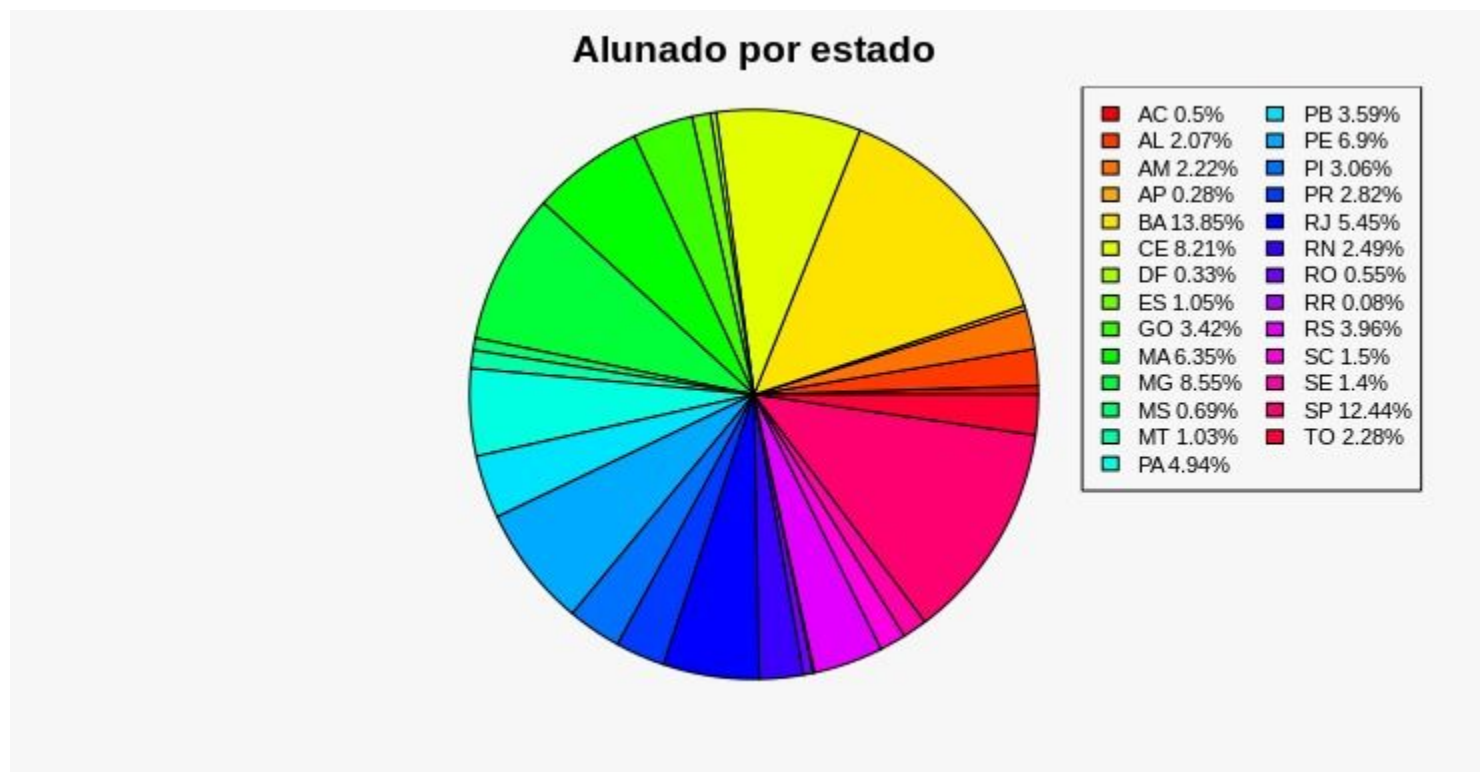
Conforme as análises pode-se notar que a Região que mais possui alunos atendidos pelo programa é o Nordeste, seguido de Sudeste, Norte, Sul e Centro Oeste respectivamente.



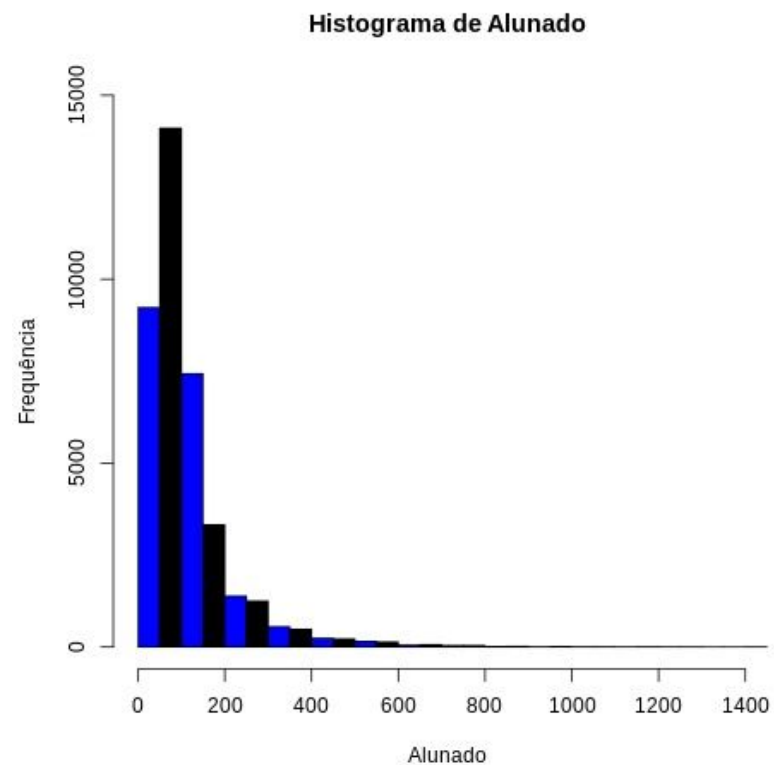
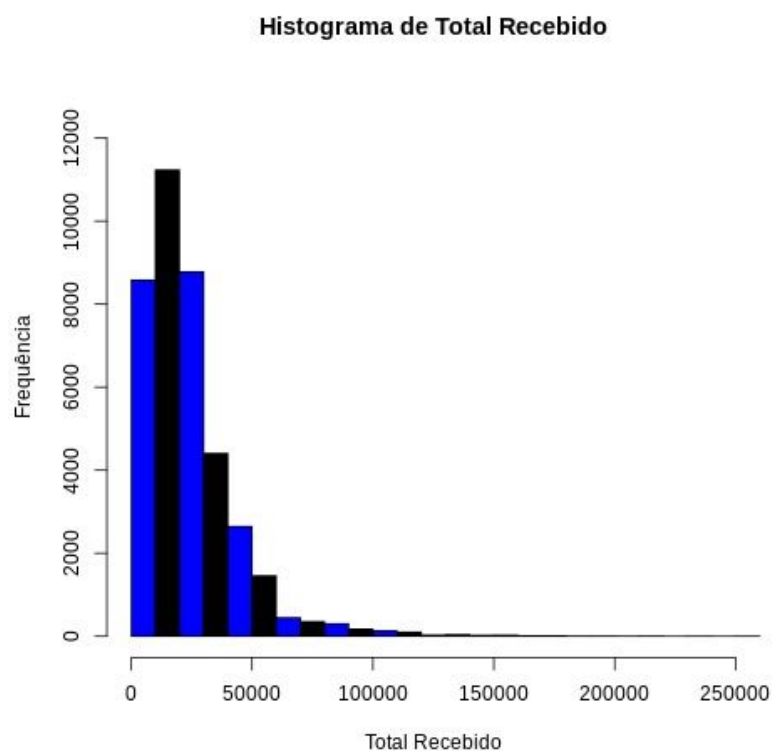
O total recebido por região possui uma proporção parecida com a de alunado apresentado anteriormente. A região que mais recebe investimentos é o Nordeste, seguido de Sudeste, Norte, Sul e Centro Oeste respectivamente.



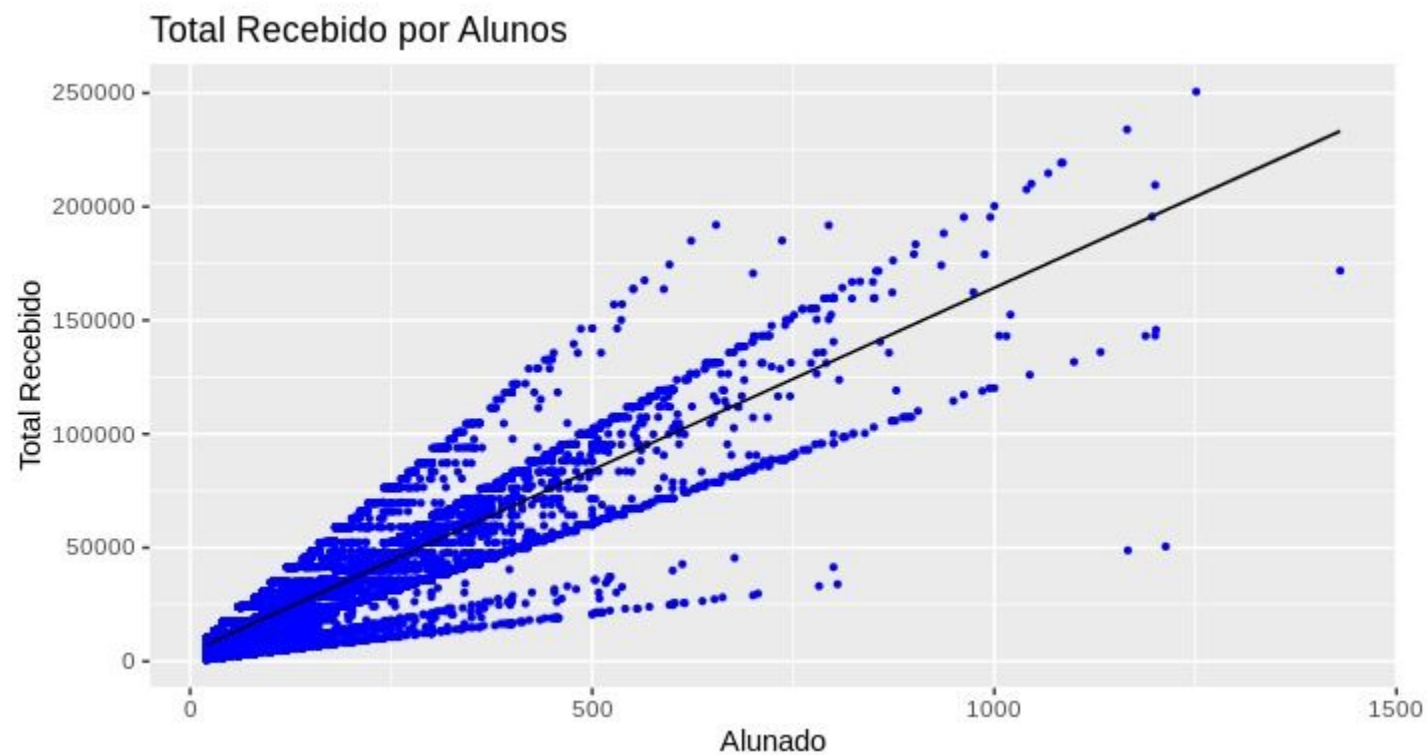
Separando o alunado por estado pode-se observar que o estado que mais possui alunos atendidos pelo programa é a Bahia e o estado que recebe menor investimento é Roraima.



A seguir dois histogramas são apresentados ilustrando a distribuição de frequência de alunado e total investido. Nota-se que os dois histogramas possuem um comportamento de Assimetria Positiva.



Podemos observar no gráfico que os atributos Alunado e Total Recebido possuem uma correlação considerada forte onde é possível traçar uma reta baseada na regressão linear aplicando as regras de Pearson. Nota-se também uma maior concentração dos valores na região positiva do gráfico que caracteriza uma correlação linear positiva.



Perguntas propostas

Ao olhar para nossa base de dados, os questionamentos abaixo foram surgindo e conforme os conhecimentos adquiridos no decorrer do curso foi possível explorar de forma minuciosa os dados recolhidos e informações geradas utilizando as técnicas aprendidas em sala de aula.

- Qual a quantidade total de alunos atendidos pelo programa?
 - Foram atendidos no ano de 2016, 4.659.559 alunos pelo programa Mais Educação.
- Qual escola possui quantidade maior de alunos atendidos?
 - É Colégio Estadual Santa Bernadete, localizado na Bahia com 1.430 alunos.
- Qual estado possui mais recebimentos?
 - O Estado da Bahia que recebe R\$132.752.527,00.
- Qual a média de alunos pelo atributo pela localização da Escola (Rural ou Urbana)?
 - A localização Rural possui a média de 83,02214 alunos por escola.
 - A localização Urbana possui a média de 139,76596 alunos por escola.
- Qual a média de alunos por estado?
 - AC 89,76834
 - AL 114,09005
 - AM 128,18959
 - AP 219,83051
 - BA 129,11460
 - CE 124,76060
 - DF 136,51351
 - ES 84,94271

- GO 114,19355
- MA 136,07685
- MG 104,73942
- MS 158,73529
- MT 137,06553
- PA 112,96861
- PB 108,73780
- PE 123,64923
- PI 115,38269
- PR 92,61680
- RJ 119,42501
- RN 122,12513
- RO 141,65385
- RR 117,70000
- RS 102,43531
- SC 89,09207
- SE 87,70565
- SP 143,78591
- TO 165,87520

- Qual a média dos valores recebidos por estado?

- AC R\$ 12.133,02
- AL R\$ 24.264,04
- AM R\$ 26.203,95
- AP R\$ 31.522,39
- BA R\$ 26.566,35
- CE R\$ 28.432,95
- DF R\$ 26.883,41
- ES R\$ 13.499,20
- GO R\$ 21.255,09

- MA R\$ 30.797,16
 - MG R\$ 18.761,50
 - MS R\$ 30.024,67
 - MT R\$ 24.227,91
 - PA R\$ 23.710,91
 - PB R\$ 22.781,59
 - PE R\$ 26.547,86
 - PI R\$ 24.038,17
 - PR R\$ 17.014,77
 - RJ R\$ 20.186,92
 - RN R\$ 26.022,86
 - RO R\$ 19.815,68
 - RR R\$ 8.049,73
 - RS R\$ 19.802,83
 - SC R\$ 15.528,78
 - SE R\$ 18.381,81
 - SP R\$ 21.121,85
 - TO R\$ 32.117,67
-
- O estado que mais tem alunos no programa está relacionado com o estado que mais recebe recursos?
 - O estado que mais tem alunos no programa é a Bahia com 645.573 alunos e também é o estado que mais recebe recursos com R\$ 132.831.763,00 de recebimento.

 - A quantidade de investimentos tem correlação com a localização da escola?
 - As escolas localizadas no meio rural recebem em média R\$ 289,77 por aluno, enquanto que as escolas do meio urbano recebem em média R\$ 169,04 por aluno. Considerando estes fatos as escolas rurais recebem em média mais investimento que as escolas urbanas.

- Qual região concentra o maior número de alunos atendidos pelo programa?
 - Nordeste com 2.232.701 alunos atendidos pelo programa.

Conclusão

Quando iniciamos os estudos de estatística aplicada pudemos observar alguns questionamentos feitos inicialmente, que a princípio foram perguntas básicas pois não tínhamos o conhecimento e nem imaginávamos o que era possível fazer com a estatística, porém no decorrer das aulas fomos aprimorando e melhorando os questionamentos. A estatística descritiva nos permitiu, descrever e compreender os dados da nossa base usando as medidas de tendência central, medidas de dispersão, medidas de intervalo, medidas de posição e medidas de forma. Como nossa base tem um tamanho considerável, então estas medidas foram fundamentais para observar, através de uma amostra, as informações de maneira mais simples e também nos permitiu comparar dados e concentrar as informações relevantes. Estas medidas apresentam um comportamento que pode trazer uma falsa impressão em uma análise estatística, pois por uniformizar os valores de um conjunto de dados, não representam bem os conjuntos que revelam tendências extremas, ou seja é grandemente influenciada pelos valores extremos. Por exemplo identificamos em nossa base que em Roraima existe apenas uma escola Rural quando é tirada a média geral não é possível identificar esta informação. Mas a estatística descritiva foi a ponte para a estatística inferencial onde foi possível realizar melhor a extrapolação da nossa base, neste caso através da regressão, onde observamos que a nossa base tinha uma correlação linear entre os atributos ALUNADO e TOTAL_RECEBIDO_ADESAO.

Nossa base permite ainda explorar a o atributo ESFERA_GOVERNO e MUNICIPIO_ESCOLA e tentar verificar mais a fundo a correlação entre as escolas rurais e urbanas, alunados e TOTAL_RECEBIDO_ADESAO, podendo aplicar todas as técnicas aprendidas no decorrer do curso e conseguir dar mais riqueza e detalhes as análises.