

**UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES**

BRENO DA SILVA NOGUEIRA 12400392
BRENO DE GONZALEZ BUENO 12543190
RAFAEL MACHADO DE MORAES 12702840
LUIGI TRAMONTIN DE PAULA 12727987
NICOLY G S ALFENAS 12696607

Trabalho Prático Final
Tratamento e Análise de Dados / Informações

São Paulo

2021

BRENO DA SILVA NOGUEIRA
BRENO DE GONZALEZ BUENO
RAFAEL MACHADO DE MORAES
LUIGI TRAMONTIN DE PAULA
NICOLY G S ALFENAS

Trabalho Prático Final
Tratamento e Análise de Dados / Informações

Trabalho apresentado à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para aprovação na disciplina de Tratamento e Análise de Dados / Informações.

Orientador: Prof. Regis Rossi A. Faria.

São Paulo

2021

Sumário

Resumo	5
Introdução	6
Desenvolvimento	7
Resultados	8
Conclusão	14
Referências	17

Lista de Figuras

Figura 1. Medidas de tendência central	8
Figura 2. Medidas de variabilidade	8
Figura 3. Quantis	8
Figura 4. Histograma de notas da turma A	9
Figura 5. Histograma de notas da turma B	9
Figura 6. Box-plot turma A	9
Figura 7. Box-plot turma B	10
Figura 8. Gráfico de pizza da turma A	10
Figura 9. Gráfico de pizza da turma B	10
Figura 10. Gráfico de dispersão da turma A	11
Figura 11. Gráfico de dispersão da turma B	11
Figura 12. Gráfico de dispersão das turmas A e B	12
Figura 13. Tabela de Frequência Relativa das Notas das turmas A e B	12
Figura 14. Gráfico de Dispersão da Frequência Relativa das Notas das turmas A e B com linha de tendência	13
Figura 15. Coeficiente de Correlação	13
Figura 16. Cálculo de A e B para traçar a Reta de Regressão	13

1. Resumo

Trabalho final apresentado à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para aprovação na disciplina de Tratamento e Análise de Dados / Informações, com o orientador Prof. Regis Rossi A. Faria. O objetivo do trabalho é realizar uma avaliação do desempenho das notas de prova obtidas por alunos de duas turmas diferentes da mesma matéria, sendo que a turma A fez o curso introdutório à matéria (Curso I) e a outra, a turma B, fez o curso avançado (Curso II). A partir das estatísticas obtidas, iremos apurar informações sobre as diferenças entre os perfis das turmas e as diferenças no aprendizado efetivo da matéria entre as turmas.

2. Introdução

O presente trabalho se refere ao desempenho de alunos de duas turmas que cursam a mesma matéria, porém em níveis diferentes – enquanto a turma A realiza o curso introdutório, a turma B realiza o curso avançado.

A apuração deste ofício possui a finalidade de expor características das salas examinadas a partir de dados das suas notas, o que facilitaria novas estratégias para melhorar o rendimento final das turmas.

Para a realização deste, o grupo se dividiu de forma eficiente. A integrante Nicoly tratou os dados fornecidos e gerou gráficos que facilitaram o andamento, o Breno Bueno fez a correlação entre os dados, o Breno Nogueira e o Rafael ficaram responsáveis por analisar os dados e, por fim, o Luigi redigiu a introdução e organizou o documento em geral. Finalizando todas estas etapas o grupo se juntou para escrever a conclusão do trabalho.

3. Desenvolvimento

A primeira etapa do trabalho tinha por objetivo comparar as notas de cada uma das duas turmas individualmente. Para isso foi necessário inicialmente obter as medidas de tendência central, de dispersão e os quantis a partir dos dados disponibilizados, para que fosse possível plotar gráficos e tabelas descritivas destes. Os valores encontrados podem ser observados nos resultados. (*Figura 1, Figura 2, Figura 3*)

Para a criação do primeiro gráfico ilustrativo desses dados foi utilizada a ferramenta para a criação de histogramas do Google Planilhas. Foram utilizados os valores 0 e 10 como limites dos gráficos, sendo estes a menor e a maior nota possível respectivamente. A largura de cada classe foi definida como 1,5 para que fosse possível visualizar cada uma delas, pois ao utilizar o valor 2 – que deixaria a largura das classes proporcional – o número de alunos com nota 10 deixava de ser incluído no histograma. (*Figura 4, Figura 5*)

A partir do cálculo dos quantis e limites superior e inferior também foi possível a criação de um box-plot para cada turma. (*Figura 6, Figura 7*)

Com o objetivo de mostrar de forma mais refinada as características capturadas nos dados, dois tipos de gráficos adicionais foram criados na ferramenta, sendo estes gráficos de pizza e de dispersão. (*Figura 8, Figura 9, Figura 10, Figura 11*)

Obtidos esses dados seguimos para procurar relacionar os dados das notas da turma A e B vendo se há algum tipo de relação entre eles, para isso primeiramente foi feito um gráfico de dispersão com as notas de ambas as turmas. (*Figura 12*)

Uma vez que a quantidade de alunos das salas é diferente, não é possível aplicar métodos de correlação conhecidos nos dados da forma que são apresentados, por isso para obter o coeficiente de correlação das notas das turmas utilizaremos a frequência relativa das notas, calculadas em 5 classes cada uma com um intervalo de 2 entre si. (*Figura 13*)

Com a frequência relativa podemos realizar novamente um diagrama de dispersão dos dados, dessa vez com linha de tendência. (*Figura 14*)

Se utilizando das Frequências Relativas é também possível calcular o coeficiente de correlação. (*Figura 15*)

É também possível calcular A e B para traçar a Reta de Regressão. (*Figura 16*)

4. Resultados

Figura 1. Medidas de tendência central

	Turma A	Turma B
Média	7,81	7,67
Mediana	7,50	7,87
Moda	10	9,6

Fonte: Autor.

Figura 2. Medidas de variabilidade

	Turma A	Turma B
Amplitude	10	8,4
Desvio Padrão	2,6732	2,0017
Coefc. variação	34,2423	26,1130

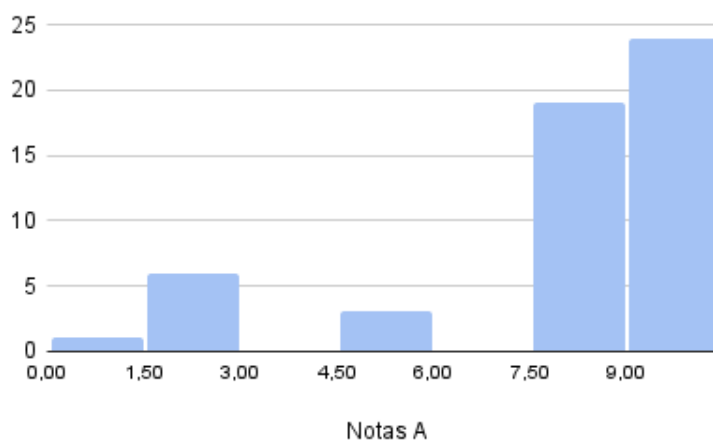
Fonte: Autor.

Figura 3. Quantis

	Turma A	Turma B
Lim. inferior	3,75	1,875
Valor min.	0	1,6
Q1	7,5	6,27
Q2	7,5	7,6
Q3	10	9,2
Valor max.	10	10
Lim. superior	13,75	13,595

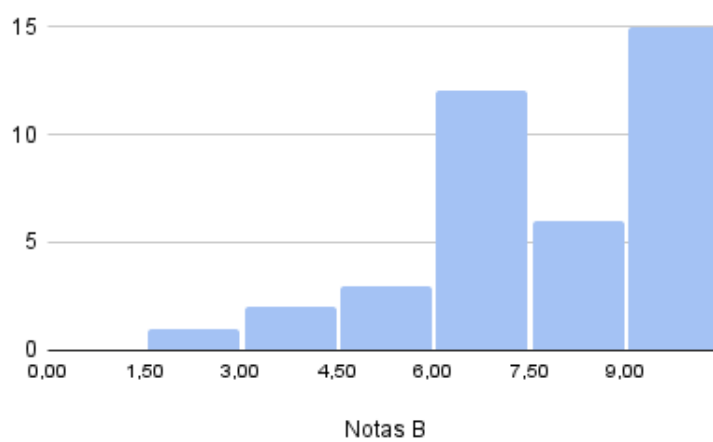
Fonte: Autor.

Figura 4. Histograma de notas da turma A



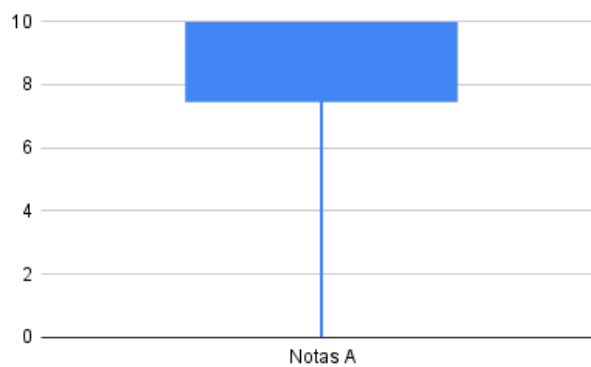
Fonte: Autor.

Figura 5. Histograma de notas da turma B



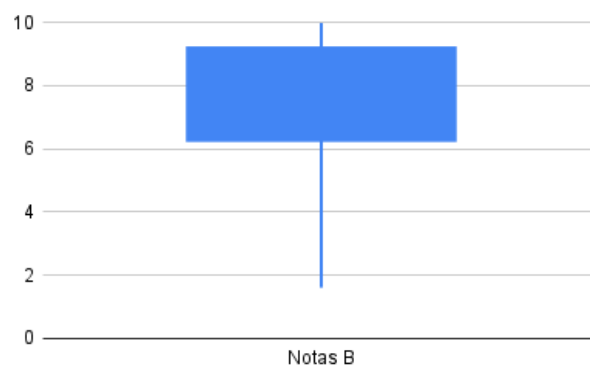
Fonte: Autor.

Figura 6. Box-plot turma A



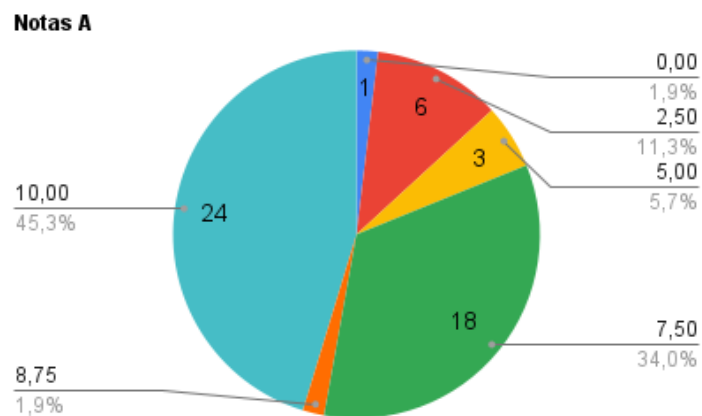
Fonte: Autor.

Figura 7. Box-plot turma B



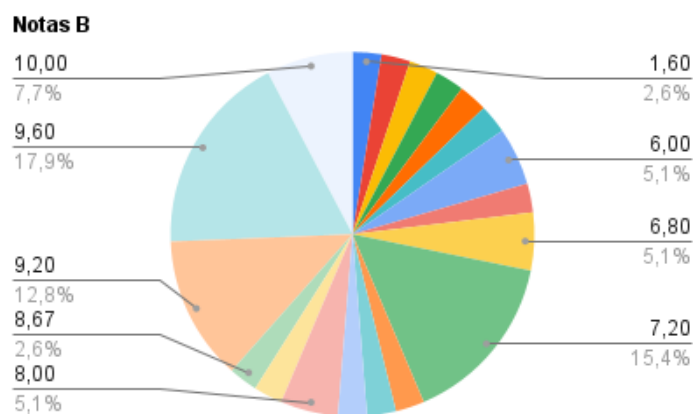
Fonte: Autor.

Figura 8. Gráfico de pizza da turma A



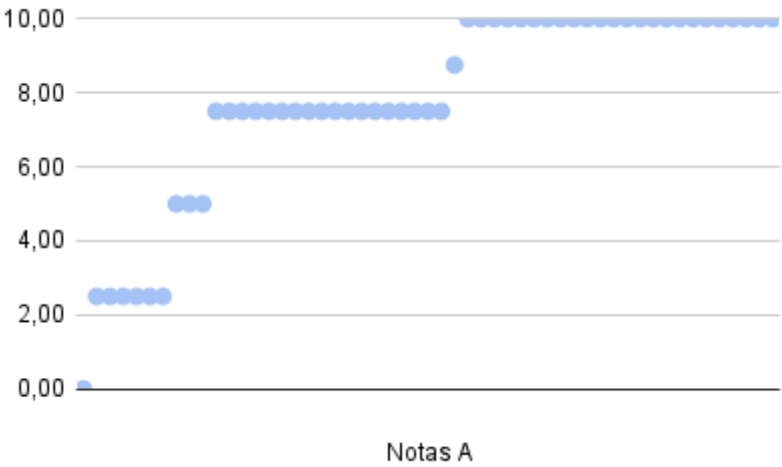
Fonte: Autor.

Figura 9. Gráfico de pizza da turma B



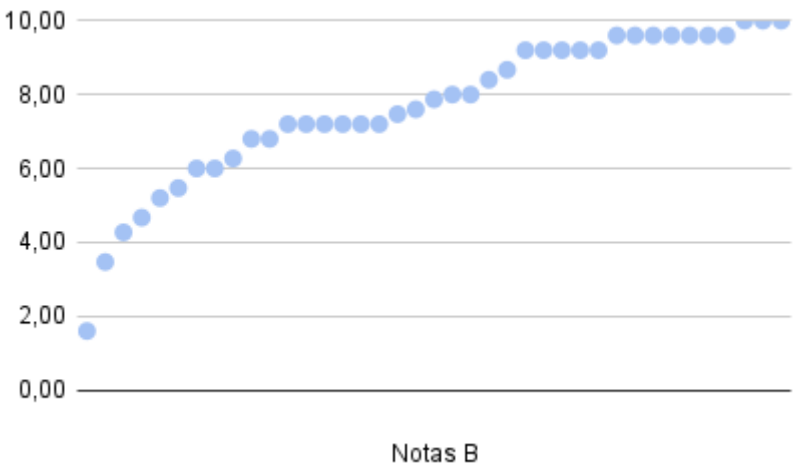
Fonte: Autor.

Figura 10. Gráfico de dispersão da turma A



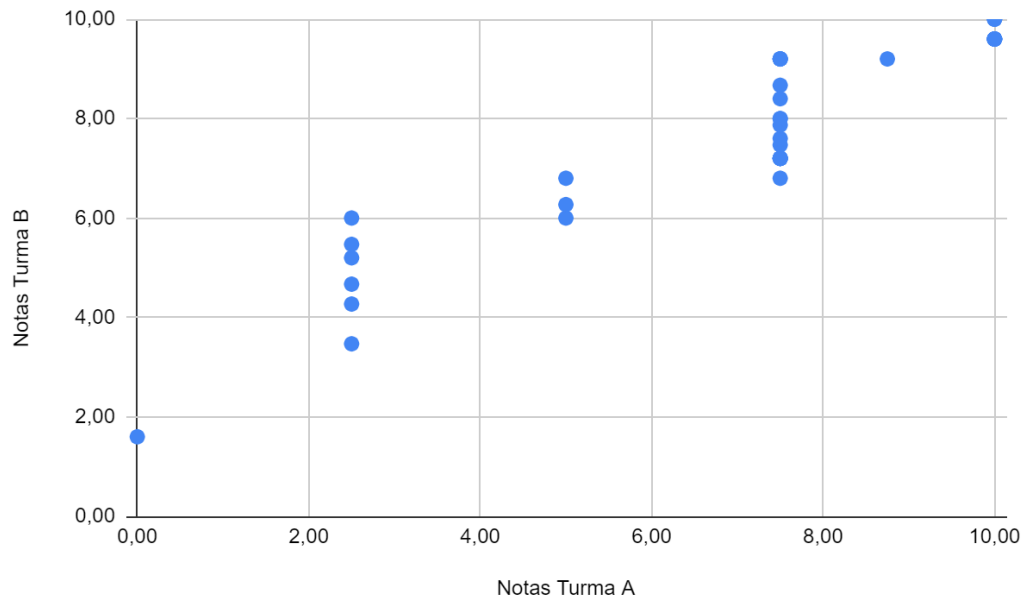
Fonte: Autor.

Figura 11. Gráfico de dispersão da turma B



Fonte: Autor.

Figura 12. Gráfico de dispersão das turmas A e B



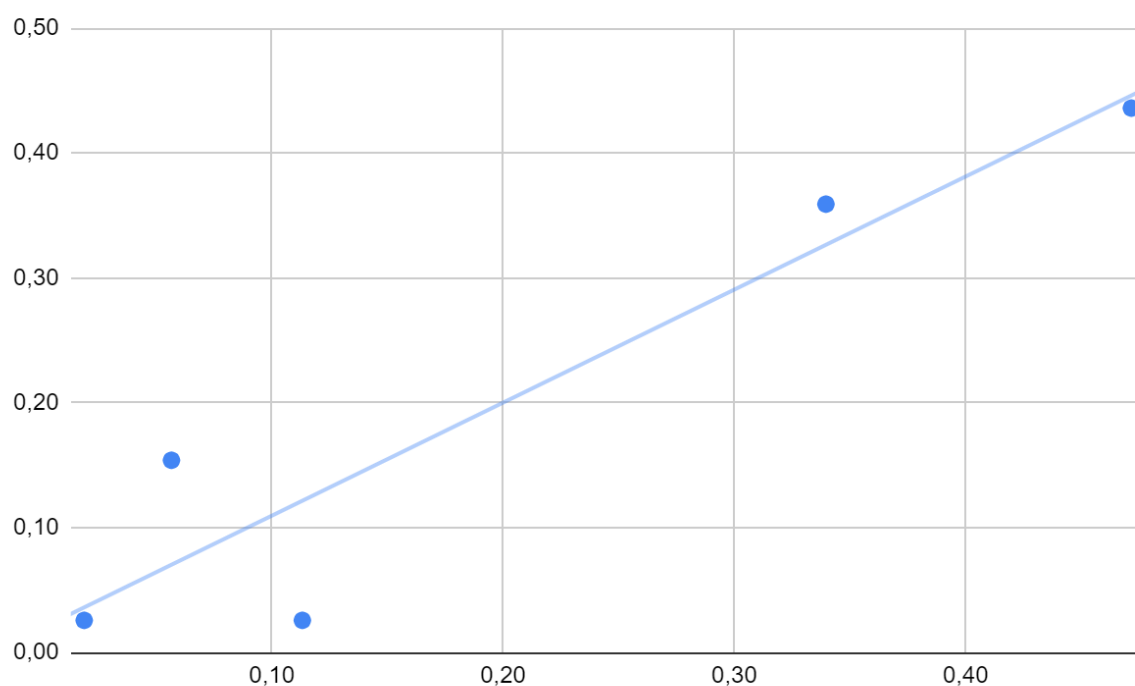
Fonte: Autor.

Figura 13. Tabela de Frequência Relativa das Notas das turmas A e B

Frequência Relativa		
Lim Sup	Turma A	Turma B
2	0,02	0,03
4	0,11	0,03
6	0,06	0,15
8	0,34	0,36
10	0,47	0,44
Total	1,00	1,00

Fonte: Autor.

Figura 14.Gráfico de Dispersão da Frequência Relativa das Notas das turmas A e B com linha de tendência



Fonte: Autor.

Figura 15.Coeficiente de Correlação

Coeficiente de Correlação = 0,94

Fonte: Autor.

Figura 16.Cálculo de A e B para traçar a Reta de Regressão

$$A = -0,02 \quad B = 1,23$$

Fonte: Autor.

5. Conclusão

Com os dados obtidos sobre as medidas resumo e os gráficos podemos traçar um perfil para cada sala, quanto ao seu desempenho, se é homogênea ou heterogênea quanto ao desempenho de seus alunos, destacando suas qualidades e características. Com essas informações em mãos podemos definir as diferenças de aprendizado de cada grupo, comparando-os para compreender melhor as diferenças.

Começando as análises com a turma A, com a média, “É a soma das observações dividido pela quantidade delas” (BUSSAB, MORETTIN, 2017), é perceptível que a maioria dos alunos passaram na prova, caso a nota de corte tenha sido 5, demonstrando que a turma como um todo aprendeu bem a matéria. Com a mediana, “É a realização que ocupa a posição central da série de observações, quando estão ordenadas em ordem crescente” (BUSSAB, MORETTIN, 2017), em 7,50, reforça a ideia obtida com a média. A moda, “é definida como a realização mais frequente do conjunto de valores observados” (BUSSAB, MORETTIN, 2017), (10) demonstrou que a maior parte da turma aprendeu a matéria de forma excelente, corroborando com a inferência feita com a média.

Indo para medidas da turma B, tivemos algumas diferenças entre os números, contudo, a diferença não foi grande o bastante para considerar a turma B deficiente no aprendizado. A média da turma B teve uma diferença de dois pontos decimais para baixo, com uma diferença tão mínima, mostra que a grande maioria dos alunos teve desempenho satisfatório, e atingiu uma nota maior que a de corte. Em contrapartida, a mediana da turma ficou em quase quatro pontos decimais acima da turma A, o que mostra que mais pessoas tiveram notas melhores na turma B, em relação ao número de pessoas da sala. Por fim, a moda, ainda que com uma diferença de quatro pontos decimais, não se distancia muito dos valores anteriores, mostrando que houve um bom aproveitamento da matéria.

Com os perfis apresentados podemos comparar as duas turmas, para assim descobrir qual turma é mais constante e homogênea, ou seja que os resultados variam menos e como se isso impacta no resultado geral da turma. Para isso vamos as medidas de dispersão, essas calculam quanto um as observações variam em relação ao todo, para essa análise será utilizado o desvio padrão, que é a raiz quadrada da variância. Esta é calculada como a média do desvio dos valores para a média aritmética, permitindo que grupos de tamanhos diferentes sejam comparados (BUSSAB, MORETTIN, 2017), algo que é muito importante para a nossa análise.

Tendo isso em vista, comparando os dois desvios padrões podemos tirar algumas conclusões sobre como uma turma se compara com a outra. A turma A teve um D.P. (desvio padrão) de 2,6732 e a turma B de 2,0017, demonstrando que mesmo as médias sendo semelhantes, a turma B é mais homogênea em seu aprendizado, com as notas variando menos. Algo que dá suporte a essa ideia é a amplitude, a diferença do menor valor para o maior valor nas observações, sendo consideravelmente maior a da turma A. Com isso concluímos que mesmo ambas tendo um desempenho geral semelhante a turma B se mostrou mais constante entre seus alunos.

Como confirmação visual dessas métricas, temos os dois gráficos principais, o histograma, e o box plot. Da turma A, o histograma nos mostra que a maior parte das notas estão no intervalo de 7,50 até 10, com furos em intervalos menores, e o box plot só confirma esse alcance, variando de aproximadamente 7,5 até 10. Já na turma B, o histograma nos diz que os dados estão mais contidos no intervalo de 6 até 10, sem muitos furos nos intervalos anteriores, que por si só já mostra uma variação mais vasta que a turma A, e o box plot refina esse intervalo com maior número de notas para aproximadamente 6,20 até 9,60, confirmando a maior amplitude da turma B.

Para os outros dois gráficos, o de pizza e o de dispersão, além de confirmar mais uma vez a maior extensão de notas da turma B, mostra um fenômeno interessante, onde a turma B possui variedade de notas muito maior, com números decimais divergindo do estático X,50 ou X,75, isso fica evidenciado no maior número de repartições do gráfico de pizza da turma B, e da curva que se forma em seu gráfico de dispersão, em contrapartida, os gráficos da turma A são mais enxutos quanto as suas divisões e diversidade de notas.

Outra forma de perceber a relação entre duas variáveis é utilizando medidas associativas, que medem o grau de correlação entre duas variáveis e sua força de regressão. Para isso usaremos o coeficiente de correlação de Pearson, uma variável que utiliza a média e o desvio padrão para determinar a relação entre duas variáveis, como as turmas possuem um número diferente de alunos inicialmente tivemos dificuldade para obter o coeficiente de correlação, porém se utilizando das frequências relativas das notas obtemos o valor de 0,94. Interpretando esse valor, percebemos que o valor é próximo de 1 demonstrando um alto grau de relação entre as variáveis.

Com essas constatações feitas, devemos olhar para o grau de confiança de nossas constatações. Como temos apenas uma métrica para manipular (a nota das duas turmas), é impossível ter uma análise muito confiável ou certa, por existirem algumas variáveis que não são explicitadas: pode ser que a nota da prova que analisamos não seja a única maneira de

avaliar os alunos, os professores podem ser diferentes, e apesar de ser explicitado, o fato de serem diferentes níveis de uma mesmo curso, contribui para que a comparação seja mais faltosa.

Além do mais, a comparação seria mais precisa se houvessem outros parâmetros para comparar à nota, com a intenção de achar mais associações que pudessem caracterizar mais a fundo as duas turmas, como por exemplo: tempo de estudo, tempo e quantidade de aulas de cada turma, quantidade de atividades feitas como preparação para a prova, entre outras variáveis.

O maior desafio encontrado nesse projeto foi para gerar e interpretar o gráfico com a reta de regressão (*Figura 14*). Isso foi causado pela quantidade de dados ser diferente entre a turma A e a turma B, dificultando a geração do gráfico. Para resolver esse problema optamos por utilizar classes e seus limites, assim possibilitando ter algo que pudesse ser interpretado, com os seus limites.

Em suma, podemos vagamente dizer que as duas turmas possuem desempenhos semelhantes, com um bom aproveitamento da matéria no contexto da única prova que tiveram.

6. Referências

MORETTIN, Pedro Alberto; BUSSAB, Wilton Oliveira. **Estatística básica**. Saraiva Educação SA, 2017.