PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS NÚCLEO DE EDUCAÇÃO A DISTÂNCIA

Pós-graduação *Lato Sensu* em Ciência de Dados e Big Data

Pablo Teixeira de Souza

OBSERVAÇÃO DOS IMPACTOS NAS OCORRÊNCIAS AERONÁUTICAS NO BRASIL A PARTIR DE RECOMENDAÇÕES DE SEGURANÇA

Pablo Teixeira de Souza

OBSERVAÇÃO DOS IMPACTOS NAS OCORRÊNCIAS AERONÁUTICAS NO BRASIL A PARTIR DE RECOMENDAÇÕES DE SEGURANÇA

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Ciência de Dados e Big Data como requisito parcial à obtenção do título de especialista.

Belo Horizonte 2020

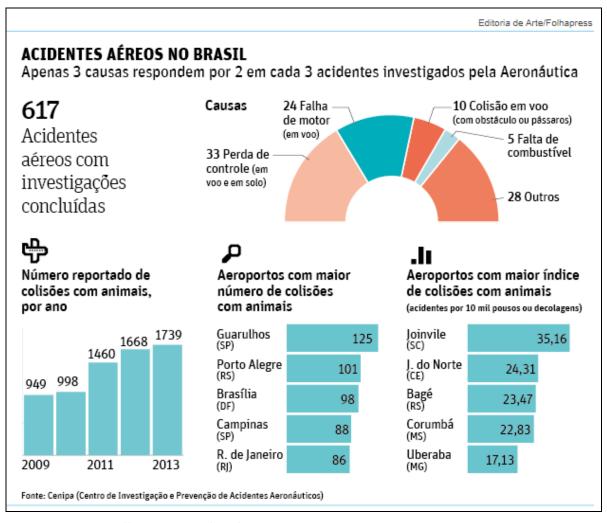
SUMÁRIO

1. Introdução	4
1.1. Contextualização	4
1.2. O problema proposto	6
2. Coleta de Dados	6
3. Processamento/Tratamento de Dados	8
4. Análise e Exploração dos Dados	11
5. Apresentação dos Resultados	14
6. Links	17
REFERÊNCIAS	18

1. Introdução

1.1. Contextualização

A atualização de dados sobre ocorrências aeronáuticas pelo CENIPA - Centro de Investigação e Prevenção de Acidentes Aeronáuticos da Força Aérea Brasileira, em 2020 mostra que na última década o fator humano ainda é o maior causador de incidentes, incidentes graves ou acidentes. Sites como "O GLOBO" e "FOLHA" compilam estes dados em suas matérias e exibem gráficos agrupados sobre causa de falha, aeroportos com maiores índices, investigações concluídas etc., atém cruzarem dados sobre a investigação de ocorrências com vítimas fatais ou de repercussão popular.



 $\textbf{Fonte:}\ https://oglobo.globo.com/brasil/falha-humana-o-fator-que-mais-contribui-para-acidentes-aereos-no-brasil-13683583$





Morte de Campos: Investigações do Cenipa levam em média mais de mil dias

Análise de fatos isolados pode levar a conclusões precipitadas, diz FAB

Aeronáutica afirmou em nota que não trabalha com a causa do acidente que matou Eduardo Campos



Principal suspeita da FAB para acidente aéreo é de falha humana



Cenipa não é mais obrigado a compartilhar apuração de quedas de aviões com civis

Fonte: https://www1.folha.uol.com.br/poder/2014/08/1504291-perda-de-controle-da-aeronave-e-uma-das-principais-causas-de-acidente.shtml

O CENIPA atua principalmente através de seus sistemas de RCSV - Relato ao CENIPA para a Segurança de Voo, que é uma ferramenta de prevenção do Sistema de Investigação e Prevenção de Acidentes Aeronáuticos (SIPAER), cuja finalidade é relatar ao Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA) uma situação com potencial de risco para a segurança de voo. FOD - Foreign Object Damagee, que relata danos por objeto estranho e pelo RELPREV Relato de Prevenção, documento formal destinado ao reporte voluntário de uma situação de perigo, real ou potencial, ou que delas teve conhecimento, facilitando a identificação reativa e proativa dos perigos à segurança operacional.

"Com o objetivo de mitigar o risco de ocorrências, o CENIPA atua por meio dos sete Serviços Regionais de Investigação e Prevenção de Acidentes Aeronáuticos (SERIPA), espalhados pelo território brasileiro"

(CENIPA, http://www2.fab.mil.br/cenipa/index.php/prevencao-de-acidentes)

Todo incidente, incidente grave ou acidente deve ser registrado no CENIPA, que por sua vez fará uma investigação, podendo ou não gerar recomendações de segurança e/ou obrigações de alterações de procedimentos, troca de componentes ou paralisação de aeronaves, aeroportos ou empresas.

O objetivo desta análise é a verificação da obtenção de impactos positivos nas ocorrências, ou seja, relacionar a boa prática de seguir recomendações e a diminuição de ocorrências.

1.2. O problema proposto

Verificar a obtenção de impactos positivos nas ocorrências aeronáuticas no Brasil a partir de recomendações de segurança geradas pelo CENIPA, através de dados públicos e atualizados em 2020. Esta análise tem o objetivo de relacionar a boa prática de seguir estas recomendações de segurança a diminuição de ocorrências aeronáuticas do mesmo tipo.

(Why?) Este problema é importante pois, caso comprovada a relação direta, é possível aumentar o potencial de prevenção de incidentes, incidentes graves ou acidentes aeronáuticos.

(Who?) Os dados analisados são do CENIPA – Centro de Investigação e Prevenção de Acidentes Aeronáuticos, da Força Aérea Brasileira.

(What?): OS objetivos são identificar os maiores fatores contribuintes, seus subtipos, a relação entre as ocorrências e as recomendações de segurança.

(Where?): Os dados disponibilizados são referentes os 27 estados mais o Distrito Federal do Brasil.

(When?): Estão sendo analisados dados entre os anos e 2008 a 2017.

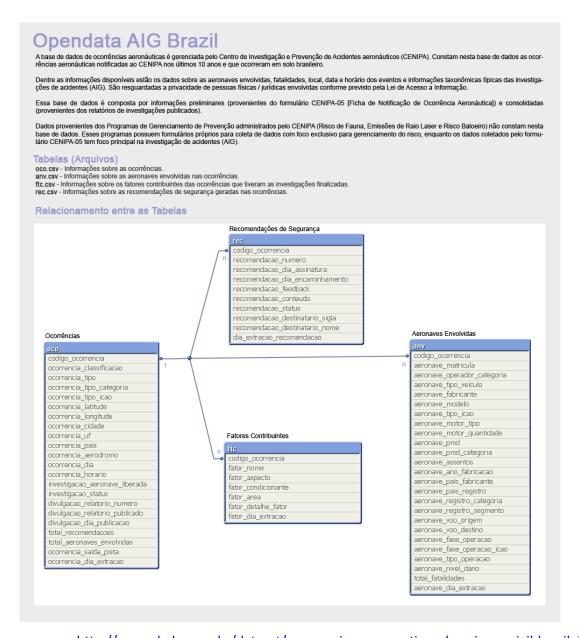
2. Coleta de Dados

Os dados utilizados nesta análise são públicos, atualizados em 2020 e disponibilizados pelo CENIPA, obtidos em 01 de fevereiro de 2020 através do link http://www.dados.gov.br/dataset/ocorrencias-aeronauticas-da-aviacao-civil-brasileira

Foram disponibilizados em formato csv, com a primeira linha sendo o cabeçalho, com suas colunas delimitadas pelo caractere "~" e code de origem "UTF 8". Todos os arquivos foram importados para arquivo no formato xlsx.

São 4 tabelas contendo as ocorrências aeronáuticas (arquivo oco.csv gerou o arquivo ocorrencias.xlsx), os fatores contribuintes para as ocorrências aeronáuticas (arquivo ftc.csv gerou o arquivo fatorescontribuintes.xlsx), as recomendações de segurança geradas após as ocorrências aeronáuticas (arquivo rec.csv gerou o arquivo recomendações.xlsx) e as aeronaves envolvidas nas ocorrências aeronáuticas (arquivo anv.csv gerou o arquivo aeronavesenvolvidas.xlsx).

O modelo abaixo demostra o relacionamento entre a tabela de ocorrências, Recomendações de Segurança, Fatores Contribuintes e aeronaves envolvidas através do atributo "codigo_ocorrencia". O relacionamento é sempre de um para muitos a partir da tabela de ocorrências.



Para a análise foram criados 4 datasets principais representando cada uma destas tabelas: oco – populado com a tabela de ocorrências, rec – populado com a tabela de recomendações de segurança, fat – populado com a tabela de fatores contribuintes e era – populado com a tabela de aeronaves envolvidas.

```
# Lê Ocorrencias
oco <-read_excel("D:/Pablo/PUC - Especialização Ciência de Dados e Big Data/13. TCC - Ciência de Dados e Big Data/Ocorrencias.xlsx")
#View(oco)
attach(oco)

# Lê Aeronaves Envolvidas
aer <-read_excel("D:/Pablo/PUC - Especialização Ciência de Dados e Big Data/13. TCC - Ciência de Dados e Big Data/AeronavesEnvolvidas.xlsx")
aer <a href="mailto:aeroaque.ano_fabricacao_idade">aer <a href="mailto:aeroaque.ano_fabricacao_idade</a> <- (2020 - aer saeronave_ano_fabricacao)
#View(aer)
attach(aer)

# Lê Fatores Contribuintes
fat <-read_excel("D:/Pablo/PUC - Especialização Ciência de Dados e Big Data/13. TCC - Ciência de Dados e Big Data/FatoresContribuintes.xlsx")
#View(fat)
attach(fat)
# Lê Recomendacoes
rec <-read_excel("D:/Pablo/PUC - Especialização Ciência de Dados e Big Data/13. TCC - Ciência de Dados e Big Data/Recomendacoes.xlsx")
#View(frec)
#View(frec)
#US Recomendacoes
#US Recomenda
```

3. Processamento/Tratamento de Dados

O dataset oco foi populado com 5461 registros e não contém duplicações, a leitura foi feita diretamente do arquivo Ocorrencias.xlx e possui os seguintes atributos:

Nome	Tipo	Descrição
codigo_ocorrencia	Texto	Código Identificador Ocorrência
ocorrencia_classificacao	Texto	Classificação Principal da Ocorrência
ocorrencia_tipo	Texto	Tipo da Ocorrência
ocorrencia_tipo_categoria	Texto	Categoria da Ocorrência
ocorrencia_tipo_icao	Texto	Código ICAO da Ocorrência (1)
ocorrencia_latitude	Número	Latitude Georeferenciada da Ocorrência
ocorrencia_longitude	Número	Longiture Georeferenciada da Ocorrência
ocorrencia_cidade	Texto	Cidade de Registro da Ocorrência
ocorrencia_uf	Texto	Unidade Federativa de Registro da Ocorrência
ocorrencia_pais	Texto	País de Registro da Ocorrência
ocorrencia_aerodromo	Texto	Código do Aeroclube ou Aeroporto de Registro da Ocorrência
ocorrencia_dia	Texto	Data da Ocorrência
ocorrencia_ano	Número	Ano da Ocorrência
ocorrencia_horario	Texto	Hora de Registro da Ocorrência
investigacao_aeronave_liberada	Texto	Indicador de Liberação de Investigação da Aeronave
investigacao_status	Texto	Status da Investigação
divulgacao_relatorio_numero	Texto	Número da Investigação
divulgacao_relatorio_publicado	Texto	Indicador de Divulgação da Investigação

divulgacao_dia_publicacao	Texto	Data da Publicação da Investigação
total_recomendacoes	Número	Quantidade de Recomendações
total_aeronaves_envolvidas	Número	Quantidade de Aeronaves Envolvidas
ocorrencia_saida_pista	Texto	Indicador de Ocorrência em Saída de Pista
ocorrencia_dia_extracao	Texto	Data da Extração do Registro

(1) O código aeroportuário ICAO é um código composto por quatro letras que designa aeroportos em todo o mundo, contando com 180 países membros. É definido pela Organização da Aviação Civil

O dataset fat foi populado com 4631 registros e não contém duplicações, a leitura foi feita diretamente do arquivo FatoresContribuintes.xlx e possui os seguintes atributos:

Nome	Tipo	Descrição
codigo_ocorrencia	Texto	Código Identificador Ocorrência
recomendacao_numero	Texto	Código Identificar da Recomendação
recomendacao_dia_assinatura	Texto	Data da Assinatura da Recomendação
recomendacao_dia_encaminhamento	Texto	Data do Encaminhamento da Recomendação
recomendacao_feedback	Texto	Data do Feedback da Recomendação
recomendacao_conteudo	Número	Descrição da Recomendação
recomendacao_status	Número	Status da Recomendação
		Sigla da Agência, Órgão ou Outro Destinatário da
recomendacao_destinatario_sigla	Texto	Recomendação
		Nome da Agência, Órgão ou Outro Destinatário da
recomendacao_destinatario_nome	Texto	Recomendação
dia_extracao_recomendacao	Texto	Data da Extração do Registro

O dataset rec foi populado com 2060 registros e não contém duplicações, a leitura foi feita diretamente do arquivo Recomendacoes.xlx e possui os seguintes atributos:

Nome	Tipo	Descrição
codigo_ocorrencia	Texto	Código Identificador Ocorrência
fator_nome	Texto	Nome do Fator Contribuinte
fator_aspecto	Texto	Aspecto do Fator Contribuinte
fator_condicionante	Texto	Condicionante do Fator Contribuinte
fator_area	Texto	Área do Fator Contribuinte
fator_detalhe_fator	Número	Descrição do Fator Contribuinte
fator_dia_extracao	Número	Data da Extração do Registro

O dataset era foi populado com 5524 registros e não contém duplicações, a leitura foi feita diretamente do arquivo AeronavesEnvolvidas.xlx e possui os seguintes atributos:

Nome	Tipo	Descrição
codigo_ocorrencia	Texto	Código Identificador Ocorrência
aeronave_matricula	Texto	Matrícula Identificadora da Aeronave
aeronave_operador_categoria	Texto	Categoria do Operador da Aeronave
aeronave_tipo_veiculo	Texto	Tipo de Aeronave
aeronave_fabricante	Texto	Fabricante da Aeronave
aeronave_modelo	Texto	Modelo da Aeronave
aeronave_tipo_icao	Texto	Código ICAO da Aeronave (1)
aeronave_motor_tipo	Texto	Tipo de Motor da Aeronave
aeronave_motor_quantidade	Texto	Classificação de Quantidade de Motores da Aeronave
aeronave_pmd	Número	Peso Máximo de Decolagem da Aeronave
aeronave_pmd_categoria	Texto	Range de Peso Máximo de Decolagem da Aeronave
aeronave_assentos	Número	Quantidade de Assentos da Aeronave
aeronave_ano_fabricacao	Número	Ano de Fabricação da Aeronave
aeronave_pais_fabricante	Texto	País Fabricante da Aeronave
aeronave_pais_registro	Texto	País de Registro da Aeronave
aeronave_registro_categoria	Texto	Categoria da Aeronave
aeronave_registro_segmento	Texto	Segmento de Registro da Aeronave
aeronave_voo_origem	Texto	Aeródromo ou Aeroporto de Origem do Voo
aeronave_voo_destino	Texto	Aeródromo ou Aeroporto de Destino do Voo
aeronave_fase_operacao	Texto	Fase de Operação da Aeronave na Ocorrência
aeronave_fase_operacao_icao	Texto	Código ICAO da Fase de Operação na Ocorrência (2)
aeronave_tipo_operacao	Texto	Tipo de Operação da Aeronave
aeronave_nivel_dano	Texto	Identificador de Tipo de Dano na Aeronave causado pela Ocorrência
total_fatalidades	Número	Total de Vítimas fatais contabilizadas diretamente Ligadas a Ocorrência
aeronave_dia_extracao	Texto	Data da Extração do Registro

- (1) O código de aeronave ICAO é um código que designa tipos de aeronaves em todo o mundo, contando com 180 países membros. É definido pela Organização da Aviação Civil
- (2) O código da Fase de Operação ICAO é um código que designa a fase de operação durante uma ocorrência em todo o mundo, contando com 180 países membros. É definido pela Organização da Aviação Civil

Não foi necessário o tratamento direto de nenhum tipo de dado, uma vez que os mesmos já se encontram tratados para as análises objetos deste trabalho. Os filtros, condições e relacionamentos foram tratados nos datasets que serão demonstrados no decorrer deste trabalho.

4. Análise e Exploração dos Dados

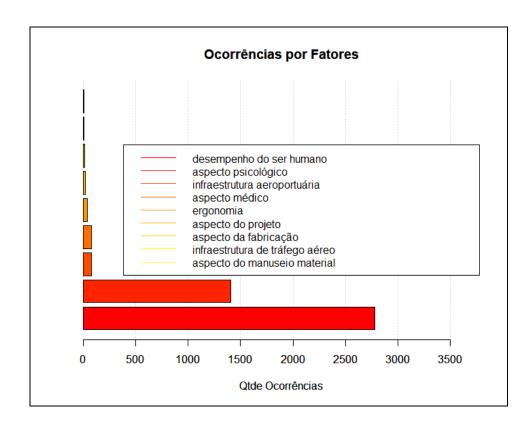
Para a realização de análise exploratória, foi realizado inicialmente a criação do dataset tot_fatotores que agrupa os aspectos dos fatores contribuintes. Foi feita uma consulta direta do dataset fat. O objetivo é observar de forma gráfica os maiores aspectos ofensores.

Foi observado que das 5461 ocorrências, categorizadas em 9 tipos, totalizando 4631 fatores contribuintes, 2782 ocorrências tiveram como causa o Desempenho Humano, representando 50,84% do total de ocorrências. Também foi observado que 1408 ocorrências tiveram como causa o aspecto psicológico, representando 25,73 do total da amostra. Considerando que estas duas causas tem sua raiz no aspecto humano, a soma destes dois ofensores é de 4190, ou 75,57% do total da amostra. As outras 1271 ocorrências, representando 23,43% desta amostra se dividem em aspecto da fabricação, aspecto do manuseio do material, aspecto do projeto, aspecto médico, ergonomia, infraestrutura aeroportuária, infraestrutura de tráfego aéreo e outros.

A hipótese é que mais de 75% das ocorrências aeroportuárias com sua origem em fatores humanos, portanto, devem ter tratamento específico para este tipo de causa raiz.

"fatores humanos" é uma expressão que ainda há de denir-se claramente, dado que quando tais palavras são utilizadas na linguagem cotidiana normalmente se referem a qualquer fator relacionado aos seres humanos. O elemento humano é a parte mais -exível, adaptável e valiosa do sistema aeronáutico, mas é também a mais vulnerável a influências que podem afetar negativamente seu comportamento"

Fonte: https://www2.anac.gov.br/arquivos/carta/fatores hum ok.pdf

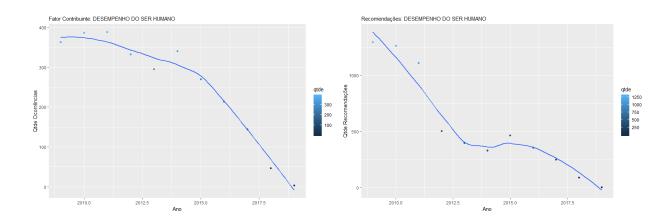


Após a primeira observação foi realizado um estudo mais aprofundado sobre a correlação entre o número de ocorrências e as recomendações feitas após estas mesmas ocorrências, categorizadas pelos 2 maiores ofensores, desempenho do ser humano e aspecto psicológico com o objetivo de verificar a efetividade desta ação.

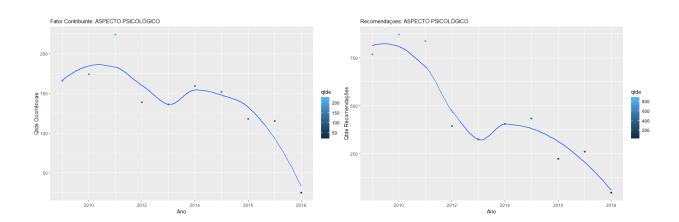
Para esta análise foi realizado um merge entre os datasets oco, rec e fat com o objetivo de relacioná-los tendo como código identificador o atributo codigo_ocorrencia, em seguida foram criados 2 novos datasets, cada um para um ofensor diferente.

Foi observado que para causas de desempenho do ser humano, a base histórica mostra uma queda acentuada a partir de 2015, quando ao mesmo tempo há estabilidade nas recomendações relacionadas ao mesmo tipo de causa. Para os

anos seguintes, pode observar que as recomendações caem em maior proporção se comparadas as ocorrências.



Para as causas de aspecto psicológico, a base histórica mostra uma harmonia entre a quantidade de ocorrências e as recomendações. Porém é possível observar que nos anos de 2008 e 2009 há um número maior de recomendações.



A hipótese é que a medida em que ocorrem incidentes, incidentes graves ou acidentes aeroportuários, há a adoção de medidas específicas e até mesmo preventivas em relação ao fator humano e que após a adoção destas recomendações as ocorrências diminuem, consequentemente diminuindo a necessidade de recomendações.

5. Apresentação dos Resultados

Foi utilizado o modelo de regressão linear de Pearson (r) para verificar a correlação entre a quantidade de recomendações realizadas e as ocorrências subsequentes registradas.

"O coeficiente de correlação de Pearson (r) ou coeficiente de correlação produto-momento ou o r de Pearson mede o grau da correlação linear entre duas variáveis quantitativas. É um índice adimensional com valores situados ente -1,0 e 1.0 inclusive, que reflete a intensidade de uma relação linear entre dois conjuntos de dados. Este coeficiente, normalmente representado pela letra "r" assume apenas valores entre -1 e 1. r= 1 Significa uma correlação perfeita positiva entre as duas variáveis. r= -1 Significa uma correlação negativa perfeita entre as duas variáveis - Isto é, se uma aumenta, a outra sempre diminui. r= 0 Significa que as duas variáveis não dependem linearmente uma da outra. No entanto, pode existir uma outra dependência que seja "não linear". Assim, o resultado r=0 deve ser investigado por outros meios."

Fonte: http://www.inf.ufsc.br/~vera.carmo/Correlacao/Correlacao Pearson Spearman Kendall.pdf

Para o ofensor desempenho do ser humano que representa 50,84% do total amostral, foi possível identificar através correlação de Pearson, o resultado de 99,58%. O valor de p foi de 1,06 é considerado uma correlação muito forte. O intervalo de confiança foi entre 98,3% e 99,8%, portanto não desvio da média sobre o valor amostral. Os valores residuais estão próximos de 1 e não apresentaram outlines. Após o ajuste o valor do Intercept não foi significativo ao nível de 5% e foi significativo ao nível de 10%, obtendo como resultado o valor de 6,29%.

```
r1 <- cor.test(dis_rec_fat_DSH$qtde,rec_fat_DSH$TotRecomendacoes)
r1

Pearson's product-moment correlation

data: dis_rec_fat_DSH$qtde and rec_fat_DSH$TotRecomendacoes
t = 32.994, df = 9, p-value = 1.063e-10
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.9836735 0.9989709
sample estimates:
cor
0.9958916
```

```
Coefficients:

Estimate Std. Error t value Pr(>|t|)

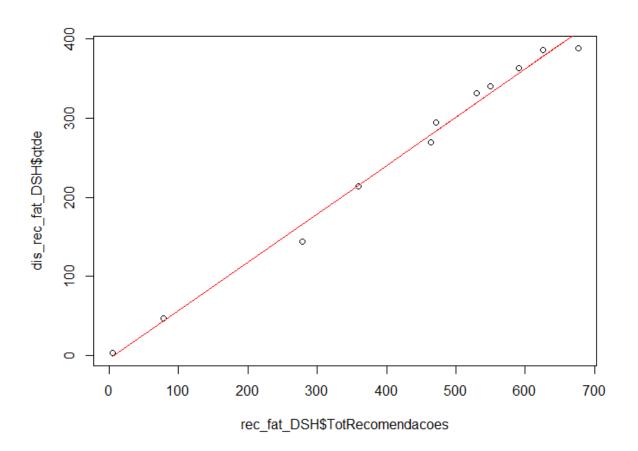
(Intercept) -4.35857 8.71166 -0.50 0.629

rec_fat_DSH$TotRecomendacoes 0.61135 0.01853 32.99 1.06e-10 ***

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Após estas observações é possível observar que para o ofensor desempenho do ser humano, a correlação é positiva, ou seja, ao se fazer recomendações de segurança, pode ser observado a diminuição de incidentes, incidentes graves ou acidentes aeroportuários.



Para o ofensor aspectos psicológicos que representa 25,73% do total amostral, foi possível identificar através correlação de Pearson, o resultado de 85,94%. O valor de p foi de 0,01 é considerado uma correlação muito fraca. O intervalo de confiança foi entre 50,07% e 96,62%, havendo, portanto, desvio da média sobre o valor amostral. Os valores residuais não estão próximos de 1 e apresentaram outlines. Após o ajuste o valor do Intercept não foi significativo ao

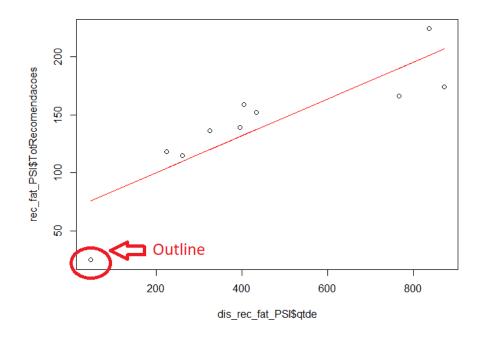
nível de 5% e não foi significativo ao nível de 10%, obtendo como resultado o valor de 0,04%.

```
r2 <- cor.test(rec_fat_PSI$TotRecomendacoes, dis_rec_fat_PSI$qtde)
r2
```

```
Pearson's product-moment correlation

data: rec_fat_PSI$TotRecomendacoes and dis_rec_fat_PSI$qtde
t = 4.7544, df = 8, p-value = 0.001437
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.5007640 0.9662146
sample estimates:
cor
0.8594201
```

Após estas observações é possível observar que para o ofensor aspectos psicológicos, não possível através do modelo de Pearson que exista uma correlação entre ao se fazer recomendações de segurança, pode ser observado a diminuição de incidentes, incidentes graves ou acidentes aeroportuários.



6. Links

Vídeo: https://youtu.be/6RKsVWUPqFQ

Scripts e demais Arquivos: https://github.com/pabloteixeiradesouza/TCC-PUC

REFERÊNCIAS

CENIPA – CENTRO DE INVESTIGAÇÃO E PREVENÇÃODE ACIDENTES AERONÁUTICOS. **PORTAL BRASILEIRO DE DADOS ABERTOS**. Disponível em: http://www.dados.gov.br/dataset/ocorrencias-aeronauticas-da-aviacao-civil-brasileira . Acesso em: 01 FEVEREIRO 2020.

CENIPA – CENTRO DE INVESTIGAÇÃO E PREVENÇÃODE ACIDENTES AERONÁUTICOS. **RSCV**. Disponível em: http://www2.fab.mil.br/cenipa/index.php/prevencao/rcsv. Acesso em: 25 FEVEREIRO 2020.

ANAC – AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL. **ANACpédia**. Disponível em: https://www2.anac.gov.br/anacpedia/por_esp/tr1070.htm. Acesso em: 25 FEVEREIRO 2020.

ANAC – AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL. **FATORES HUMANOS** Disponível em: https://www2.anac.gov.br/arquivos/carta/fatores_hum_ok.pdf. Acesso em: 21 MARÇO 2020.

GPESTATÍSTICA – COEFICIENTE DE RELAÇÃO DE PEASON Disponível em: https://gpestatistica.netlify.com/blog/correlacao/. Acesso em 21 MARÇO 2020.

CARMO.VERA – UNIVERSIDADE FEDERAL DE SANTA CATARINA Disponível em: http://www.inf.ufsc.br/. Acesso em 21 MARÇO 2020