

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Cristina Carvalho Correia Krauss

Análise dos Gastos no Pacote de Orçamento da TI Telecom

**Curitiba
2019**

Cristina Carvalho Correia Krauss

Análise dos Gastos no Pacote de Orçamento da TI Telecom

Monografia apresentada ao Programa de Especialização em *Data Science e Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Cesar Augusto Taconeli

Curitiba
2019

Análise dos Gastos no Pacote de Orçamento da TI Telecom

Analysis of expenditures in the IT Telecom budget Package

Cristina Carvalho Correia Krauss¹

¹Especialização em Data Science & Big Data, Universidade Federal do Paraná, Curitiba, PR, Brasil *

Resumo

Através do orçamento matricial foram realizadas análises dos números do OPEX que são as despesas operacionais que acontecem todos os meses para dar andamento ao negócio. Se a diferença entre o valor orçado (valor previsto) e o valor realizado (valor gasto) for muito grande ou muito pequena, para uma empresa filantrópica isso significa deixar de ajudar ao próximo. Usamos análise estatística descritiva para diferenciar quais os produtos no pacote orçamental de Tecnologia da Informação apresentam maior divergência entre custo e gasto, identificando as áreas de negócio responsáveis e suas entidades corresponsáveis. Com aplicação do modelo de regressão misto, foi possível estimar os parâmetros para descrever o comportamento médio do percentual de diferença entre valor orçado e realizado em relação produtos de TI. Os números gerados, serviram de base para repensar o processo de orçamento, a metodologia que esta sendo utilizada hoje, quais áreas necessitam de treinamento e principalmente identificar os produtos que mais apresentam discórdância no valor orçado e realizado.

Palavras-chave: orçamento matricial, modelo de regressão misto, produtos TI

Abstract

Through the matrix budget, analyses of the OPEX numbers are carried out, which are the operational expenses that take place every month to give progress to the business. If the difference between the budgeted value (predicted value) and the value realized (value spent) is too large or too small, for a philanthropic company that means no more helping others. We use descriptive statistical analysis to differentiate which products in the budget package of information technology have greater divergence between cost and expense, identifying the responsible business areas and their corresponding entities. With the application of the mixed regression model, it was possible to estimate the parameters to describe the average behavior of the percentage difference between budgeted value and performed in relation to IT products. The numbers generated, served as a basis for rethinking in the budget process, in the methodology that is being used today, which areas need training and mainly identify the products that present the most discord in the budget.

Keywords: Budget Matrix, mixed regression model, TI products

1. Introdução

O orçamento matricial é uma técnica utilizada em várias empresas, onde as despesas e receitas deixam de ser responsabilidade exclusiva do setor financeiro e passam a ser dever de várias áreas dentro da instituição, com o intuito de uma equipe ajudar a gerenciar a outra, sendo supervisionado de forma cíclica.

Orçamento é a parte de um plano financeiro estratégico que compreende a previsão de receita, valor arrecadado, e a previsão da despesa, valor gasto [1].

Um bom orçamento consegue equilibrar corretamente esses dois pontos (receita e gasto) da forma que ambos contenham números satisfatórios para o negócio. O capital utilizado para manter ou melhorar os bens físicos de uma empresa, tais como equipamentos, propriedades e imóveis, recebe a sigla de OPEX que deriva da expressão *Operational Expenditure*. As despesas operacionais OPEX são os gastos/custos contínuos para dirigir um produto, o negócio ou um sistema [2].

A tecnologia tende a ajudar as empresas na redução de custos, seja na automatização de processos, redução de mão de obra ou melhorias com novos sistemas porém para mantê-la funcionando existe um custo que

*criskrauss@gmail.com

são todos os gastos da empresa voltados para o uso da tecnologia, de suas ferramentas e sistemas.

Este trabalho tem como tema realizar uma análise dentro do orçamento do OPEX classificados como TI Telecom, subdivido em oito nichos distintos, dos anos de 2015 a 2019. Queremos entender se os valores provisionados como gastos foram devidamente realizados e quais áreas destoam dos valores planejados, orçam e não gastam, ou vice-versa, gastam e não orçam.

O problema acontece, pois, a empresa estudada possui 239 áreas diferentes, que denominamos de entidade, e cada uma dessas é responsável por gerar o seu orçamento, trabalhando com os valores do ano anterior e estimando um valor para o ano seguinte, as áreas analisam características do seu negócio e suas possíveis necessidades de gastos seguindo alguns critérios contábilísticos.

O objetivo geral está em aplicar técnicas estatísticas afim de identificar divergências nos números do Opex orçado e realizado, saber quais produtos e áreas apresentam maior discrepância dos dados e implantar um modelo de regressão para futuros orçamentos.

O estudo em questão ganha um significado maior, por se tratar de uma instituição filantrópica, onde os ganhos deixam de ser apenas melhorias nos processos financeiros e passam a ser a probabilidade de aumentar o número de pessoas a serem ajudadas.

A motivação está vinculada à missão do grupo:

"Promover a formação de cidadãos éticos, justos e solidários, por meio da educação orientada pelos valores do Evangelho, do jeito de Maria, para a transformação da sociedade"[3]

Para fazer com que os números possam auxiliar nesta tão bela missão de ajudar ao próximo foram criados conjunto de informações estatísticas, desde análise descritiva até modelos mais sofisticados buscando identificar valores não coerentes, processos a serem melhorados e áreas a serem instruídas.

Na seção 2 será abordado as técnicas que foram utilizadas, explicando os processos de construção do trabalho. Na seção 3, faremos a descrição teórica das técnicas utilizadas. Na seção 4, os resultados e conclusões obtidas. E na seção 5, será explicado como o trabalho pode ajudar a instituição e como podemos continuar com aprofundamentos nos estudos.

2. Materiais e Métodos

2.1. Orçamento Matricial

O orçamento ou análise matricial é uma metodologia gerencial para o planejamento e controle orçamentário que vem ganhando cada vez mais adeptos, principalmente pela facilidade na elaboração e pela visão cruzada e objetiva proporcionada na análise. Esta técnica está sendo usada em diversas empresas pois subdivide a responsabilidade do orçamento entre várias equipes. O modelo matricial força que cada um faça seu orçamento pensando estrategicamente como serão realizados os gastos, e avaliando se existem formas de aperfeiçoar a necessidade de recursos, quais outras formas seriam orçadas e replicadas para cada área (entidade).

Os benefícios em aderir a esse modelo são principalmente:

- O conhecimento detalhado dos gastos;
- Avaliação do desempenho individual de cada área;
- Estabelecimento de metas justas e desafiadoras;
- Melhoria da qualidade da base de dados para a tomada de decisão;
- Implementação de mudanças e melhorias contínuas no processo de gestão dos recursos;
- Elaboração do orçamento de custos que assegure o alcance de uma diretriz anual de redução de despesas.

Segue como uma matriz de gerenciamento se apresenta nas empresas:

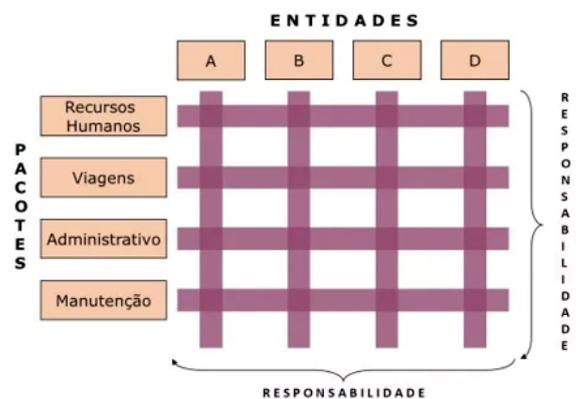


Figura 1: GMD - Orçamento Matricial

2.2. Modelo Estatístico

Um modelo estatístico, é um processo que explica o comportamento da variável resposta de acordo com variáveis explicativas ou covariáveis estatisticamente relevantes por meio de uma expressão matemática. Os estudos podem ser baseados em modelos clássicos onde se assume observações independente entre si (uma vez que os indivíduos são independentes e tem-se somente uma observação para cada um); ou então observações feitas em um mesmo individuo sejam dependentes entre si, utilizando para isso modelos que levem em consideração na sua formulação essa característica. [5]

2.3. Modelo de Regressão Linear

Análise de regressão é uma técnica estatística utilizada para investigar a relação existente entre variáveis através da construção de uma equação (um modelo). De maneira geral, essa técnica pode ser utilizada com vários objetivos, dentre os quais se pode destacar: descrever a relação entre variáveis para entender um processo ou fenômeno; prever o valor de uma variável a partir do conhecimento dos valores das outras variáveis; substituir a medição de uma variável pela observação dos valores de outras variáveis; controlar os valores de uma variável em uma faixa de interesse.

2.4. Modelo de Regressão Generalizados

Os Modelos de Regressão Linearmente Generalizados são uma extensão dos modelos de regressão simples e múltipla. Eles possibilitam utilizar outras distribuições para os erros e uma função de ligação relacionando a média da variável resposta à combinação linear das variáveis explicativas, permitindo, aumentar as suposições admitidas e examinar não somente as relações lineares entre as variáveis explicativas e a resposta. Geralmente em um problema de modelagem envolvendo essa técnica, deve-se definir:

- O comportamento (distribuição) da variável resposta;
- As variáveis explicativas;
- A função de ligação que irá ligar as variáveis explicativas às variável respostas.

Com o modelo linear generalizado é possível modelar variável de interesse que assumem a forma de contagem, contínua simétrica e assimétrica, binária e categórica. Uma das limitações desses modelos é a exigência de que os erros sejam independentes. Isso significa que eles não são capazes de modelar dados ou va-

riáveis respostas com estruturas longitudinais (amostras medidas mais de uma vez ao longo do tempo), espaciais ou multiníveis. [10]

Regressão Bernoulli / Regressão Logística

A regressão Bernoulli é utilizada na modelagem de fenômenos que podem ser resumidos em uma variável binária, ou seja, se ocorreu ou não um evento. Quando se escolhe a função de ligação 'logit', ela passa a ser chamada de regressão logística. Essa função é bastante utilizada, pois além de ser própria para dados binários auxiliando na tomada de decisão do tipo "Sim" ou "Não".

A regressão logística é muito utilizada em modelos de concessão de crédito, onde a partir de informações oferecidas pelo solicitante, a instituição financeira decide pela liberação ou não do crédito. Também é frequentemente utilizada em pesquisas clínicas que tem como objetivo verificar os fatores de influência na ocorrência ou não de uma determinada doença.

2.5. Modelo Misto

Os Modelos Lineares Generalizados Mistos ou Equações de Estimções Generalizadas são modelos lineares que apresentam somente fatores de efeitos fixos, além do erro experimental, que é sempre aleatório, são denominados modelos fixos. [4].

Um modelo misto é aquele que apresenta tanto fatores de efeitos fixos como aleatórios, além do erro experimental e da constante μ . Quando um modelo é considerado misto, sua análise de variância apresenta algumas peculiaridades, como a composição das esperanças matemáticas dos quadrados médios, cujo conhecimento permite o estabelecimento correto dos testes de hipóteses, [6]. Outro motivo de se adotar um modelo linear misto é a possibilidade de se fazer a predição de efeitos aleatórios, na presença de efeitos fixos.

2.6. Teste de AIC

A escolha do modelo apropriado, do ponto de vista estatístico, é um tópico extremamente importante na análise de dados. Busca-se o modelo mais parcimonioso, isto é, o modelo que envolva o mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento da variável resposta. [11]

2.7. Software R

O Software R é uma linguagem e ambiente para computação estatística e gráfica, sendo mais um veículo para

novos métodos em desenvolvimento de análise de dados. Tem se desenvolvido rapidamente e tem sido entendido por uma extensa coleção de pacotes. R Pacotes usados para análise deste trabalho: LMER

2.8. Conhecendo dos Dados

Para fazer com que os números possam auxiliar nesta tão bela missão de ajudar ao próximo foram criados conjunto de informações estatísticas, desde análise descritiva até modelos mais sofisticados buscando identificar valores não coerentes, processos a serem melhorados e áreas a serem instruídas.

Estamos trabalhando com uma empresa que possui muitas áreas, é possível entender essa divisão na figura abaixo, o grupo possui 12 áreas de negócio, dentro dessas áreas temos 15 Setores, ao olhar para os Setores podemos dividi-los em 69 Núcleos e estes podem ser segmentados em 239 entidades. Todas essas 239 Entidades fazem o seu orçamento



Figura 2: Divisão Estrutural.

A empresa é composta por 12 Áreas de Negócio distintas, chamadas de Frente de Missão (FM):

- **CSC** = Universidade de Santa Catarina.
- **ARC** = Auditoria, Risco e Compliance.
- **SER** = Serviços.
- **DHO** = Recursos Humanos.
- **FRS** = Financeiro.
- **MKP** = Marketing e Propaganda.
- **INF** = Infraestrutura.
- **JUR** = Área Jurídica.
- **UNV** = Universidade Paraná.
- **SOL** = Solidariedade.
- **COL** = Colégios.
- **SAD** = Hospitais.

O pacote de orçamento de TI e Telecom é composto por 8 produtos diferentes, chamados de Conta:

Aluguéis e Serviços de Manutenção de Hardware = São gastos investidos com compra de Hardware, novos computadores, aluguéis de equipamentos, estamos nos referindo a tudo que seja componente físico dentro da Tecnologia da Informação.

Licenças e Serviços de Manutenção de Software = Gastos relacionados com licenças e serviços envolvendo software para computadores, acessos a sistemas.

Materiais para Informática = Materiais relacionados a computadores, cabos, mouse, teclado, conectores para um bom funcionamento dos nossos equipamentos.

Serviços de Telefonia Fixa = Telefonia fixa de um modo geral, ramais e ligações realizadas para fora do Grupo.

Serviços de Telefonia Móvel = Telefonia Móvel, todos os telefones celulares de uso corporativo.

Serviços de TI = Gastos com fornecedores, consultorias e contratos .

Serviços de Transmissão de Dados e Internet = Gastos com Transmissão de dados , armazenamento e tráfego de dados via Internet.

Serviços e Materiais de Impressão = Gastos com tintas para impressão e papel.

Para poder seguir com as análises, utilizando o banco de dados disponibilizado, fizemos algumas escolhas / premissas para o bom andamento como por exemplo escolher uma data, para realizar o estudo. Dentre os anos de 2015 a 2019, escolhemos o ano de 2017 para base de estudo e criamos variáveis auxiliares com o intuito de ajudar na modelagem dos dados e nos modelo estatístico.

2.9. Variável Auxiliar 1

Criamos um percentual de diferença entre o OPEX orçado e realizado, foi criado para identificar a discrepância que existe entre os valores orçados pelo negócio e os valores realizados, (valores realmente gastos com as respectivas despesas). O cálculo está evidenciado na fórmula:

$$Diferença = \frac{REALIZADO\ OPEX - ORCADO\ OPEX}{ORCADO\ OPEX} \quad (1)$$

2.10. Variável Auxiliar 2

Para a aplicação do modelo de regressão linearmente generalizado misto foi criado uma variável dicotômica, classificada como 0 e 1, baseada na variável mencionada anteriormente, Quando esse percentual de diferença for entre -10% e 10% classificamos essa variável como peso 0 e caso contrário, onde tínhamos o percentual fora dos 10%, essa variável recebeu o peso 1.

A ilustração a seguir ajuda a entender a criação da variável auxiliar 2:

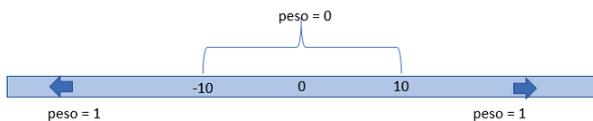


Figura 3: Criação da Variável Dicotômica.

2.11. Variável Auxiliar 3

Quando a variável composta pelo percentual de diferença entre o valor Opex Orçado e Realizado, possui uma variância muito alta, para o bom desenvolvimento do modelo, aplicamos uma função de Logaritmo neste campo, com o intuito de minimizar os erros. É possível evidenciar isso na tabela abaixo, e em seguida nos gráficos de dispersão da variável, analisar a diferença causada de uma forma que não perdemos a significância do campo.

Temos uma diferença muito grande entre o valor mínimo e o valor o máximo deste campo:

| Mínimo | 1ºQ | Mediana | Média | 3ºQ | Máximo. |
|---------|--------|---------|-------|-------|----------|
| -164.30 | -56.57 | -18.61 | 97.92 | 34.41 | 13219.99 |

Tabela 1: Demonstrando a alta variação.

A grande variação fica evidenciada também no gráfico abaixo, temos uma linha concentrada de pontos e alguns outliers que precisam ser analisados com mais cuidado, para entender o motivo. Começamos as análises pontuais por eles e suas áreas responsáveis serão acionadas para entendimento dessa divergência:

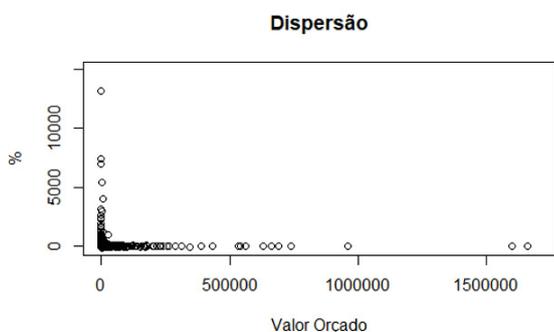


Figura 4: Dispersão.

Com a aplicação do logaritmo, os dados estão muito mais centralizados e é possível atentar que os valores

já estão mais distribuídos e homogêneos. Para continuidade dos processos estatísticos iremos utilizar a variável com a aplicação do LOG.

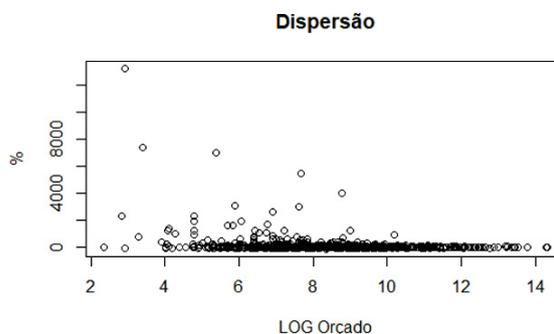


Figura 5: Dispersão - aplicação de Log.

Para desenvolver este trabalho foram utilizadas análises exploratória com técnicas gráficas e descritivas e também realizados ensaios com várias técnicas estatísticas como Modelos de Regressão Linear, Modelos Linearmente Generalizados, Modelos Mistos.

No início dos estudos não se sabe ao certo como os dados irão se comportar diante do processo de análise e da aplicação das metodologias escolhidas, como corresponderão nas técnicas aplicadas, por isso ao longo dessas análises foram aplicadas várias técnicas de modelagem, a fim de buscar o modelo ideal, isso é evidenciado por essa figura onde mostra que todos os modelos possíveis foram testados.



Figura 6: Metodologia para testes dos modelos.

Fórmula do Modelo Misto:

$$Y_{IJKL} = \alpha + Valor\ Orçado * Conta + Negócio_I + Núcleo_{IJ} + Entidade_{IJK} + \epsilon_{IJKL} \quad (2)$$

Os modelos testados foram:

$$1 + [1|Negócio] + [1|Setor] + [1|Núcleo] + [1|Entidade]$$

$$Log_{Orçado} + [1|Negócio] + [1|Setor] + [1|Núcleo] + [1|Entidade]$$

$$Log_{Orçado} + Produto + [1|Negócio] + [1|Setor] + [1|Núcleo] + [1|Entidade]$$

$$Log_{Orçado} * Produto + [1|Negócio] + [1|Setor] + [1|Núcleo] + [1|Entidade]$$

Com a variável dicotômica, pudemos testar também modelos de regressão generalizados, aonde foi possível testar 4 % diferentes dentro da variável. Este foi o modelo usado para fazer avaliação:

$$Negócio_I = N(0, \alpha^2 Negócio) \tag{3}$$

$$Núcleo_{IJ} = N(0, \alpha^2 Núcleo) \tag{4}$$

$$Entidade_{IJK} = N(0, \alpha^2 Entidade) \tag{5}$$

$$\epsilon_{IJKL} = N(0, \alpha^2 \epsilon) \tag{6}$$

$$\tag{7}$$

3. Resultados

Os resultados iniciam com uma análise estatística descritiva, explorando os números e gerando gráficos para facilitar algumas visualizações, podemos identificar que nos três últimos anos as áreas tem gerado valores superiores em seus orçamentos, ou seja, orçado mais do que realizado conforme:

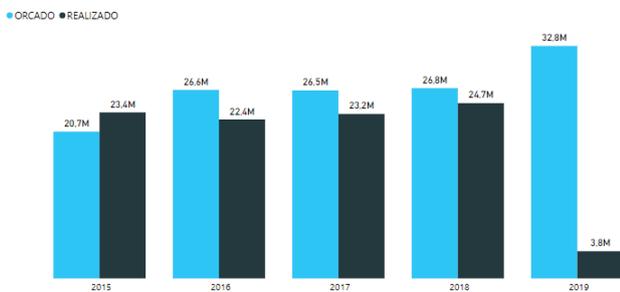


Figura 7: OPEX - Valores Orçados e Realizados de 2015 a 2019.

Nos gráficos a seguir, podemos contemplar os produtos de TI classificados em ordem de gastos. Sabendo os produtos de TI que mais ocorrem gastos, foi criada uma Squad para estudar a redução desses custos:

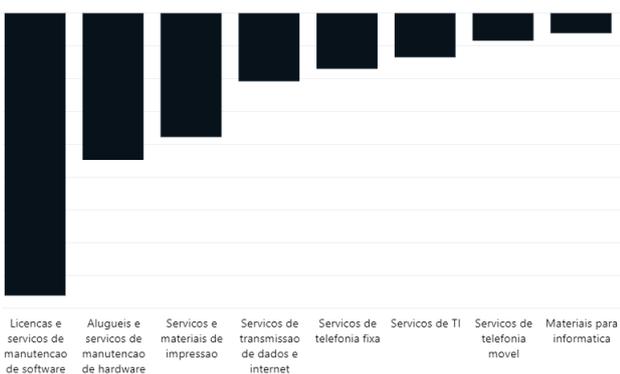


Figura 8: Produtos em ordem dos gastos.

E aplicando um modelo de visualização como o box-plot podemos ter algumas informações que completam a nossa análise. Fica evidenciado qual área e produto esta dentro da média, qual apresenta maior discrepância, podemos analisar os quartis e a mediana

dentro do grupo e na comparação de uma área com a outra.

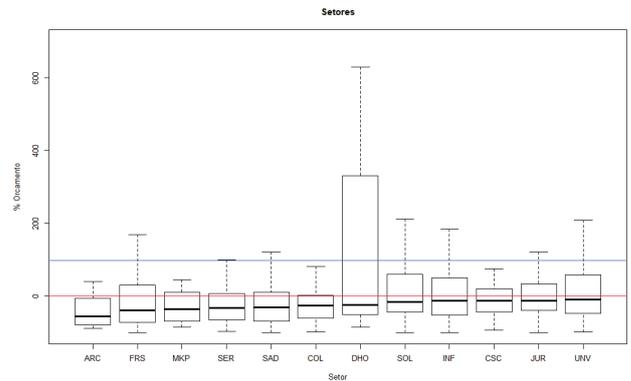


Figura 9: Valores gastos por Áreas.

1 = Área que representa maior alerta, pois apresenta maior divergência nos gastos (DHO)

2 = Área que está mais coerente em seu orçamento x realizado (UNV)

A linha vermelha do gráfico, demonstra o ponto ideal, que as equipes deveriam estar, onde a diferença entre o valor orçado e o realizado é nulo.

A linha azul do gráfico, demonstra a média, porém como temos valores muito discrepantes não é um valor a ser considerado.

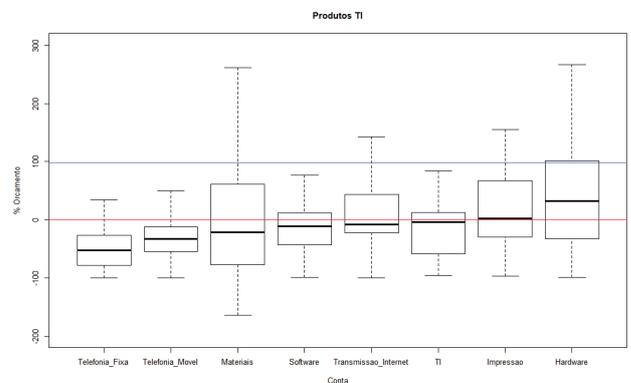


Figura 10: Valores gastos com os Produtos de TI Telecom.

1 = Produto onde os orçamentos são maiores do que os valores registrados como realizados, orçam mais que gastam

2 = Produtos onde os orçamentos são menores do que os valores registrados como realizados, gastaram mais que orçaram.

É possível identificar quais as contas apresentam maior diferença na variável do % e onde existe uma variabilidade entre as contas em relação às localidades do negócio. Pensando num exemplo prático, se fosse-

mos prover treinamento para as áreas que apresentam maior diferença, os nomes citados por nós seriam as em destaque no gráfico.

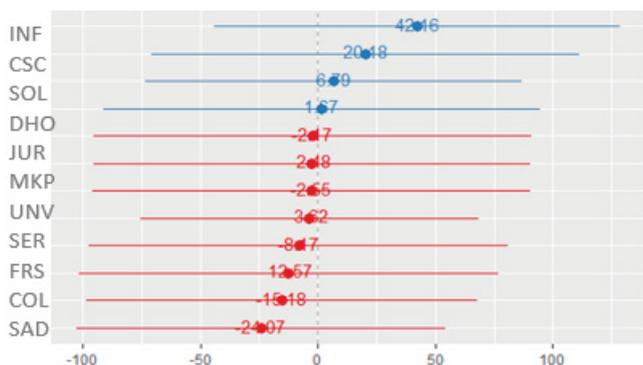


Figura 11: Variabilidade entre as Área.

Os resultados obtidos na aplicação do modelo misto, poderão ser visualizados a seguir, por serem de efeitos mistos, permitem a inclusão ao mesmo tempo de efeitos fixos e aleatórios no modelo. Observamos os efeitos aleatórios com variância zero e valores dos resíduos muito alto. Isso nos deixa claro que as variáveis Núcleo e Setor não são significativas para esse modelo. Quando encontramos diferenças estatísticas significativas entre grupos, inferimos que essas diferenças não são atribuídas ao acaso (ou ao erro, ou aos fatores não controlados), mas sim aos efeitos maiores em comum.

| Efeito | Variância | Desvio Padrão |
|----------|-----------|---------------|
| Entidade | 631976.10 | 251.388 |
| Núcleo | 0.00 | 0.00 |
| Setor | 0.00 | 0.00 |
| Negócio | 29.08 | 5.393 |
| Resíduo | 409368.88 | 639.815 |

Tabela 2: Análise do Summary, ferramenta R.

O modelo ideal é composto por:

Variável resposta = % diferença

Variáveis fixas = valor orçado , aplicado logaritmo com a variável Conta por interação

Variáveis aleatórias = Negócio e Entidade (nível do negócio mais baixo)

Seguindo o princípio de testar todos os modelos, o teste de ACI foi quem definiu o melhor modelo = modelo com menor valor e AIC é considerado o modelo de melhor ajuste:

| Modelo | AICc | D AICc | df | p-valor |
|---------|---------|--------|----|---------|
| Modelo7 | 12409.5 | 0.0 | 19 | 0.680 |
| Modelo6 | 12411.6 | 2.1 | 20 | 0.237 |
| Modelo5 | 12413.7 | 4.2 | 21 | 0.083 |
| Modelo4 | 12498.4 | 88.8 | 14 | <0.001 |
| Modelo3 | 12559.1 | 149.5 | 13 | <0.001 |
| Modelo2 | 12595.7 | 186.2 | 7 | <0.001 |
| Modelo1 | 12633.5 | 224.0 | 6 | <0.001 |

Tabela 3: Teste de AICc.

De posse do modelo foi possível alterar o efeitos fixo: conta, e perceber que as variáveis aleatórias alteravam suas saídas, verificamos como os produtos podem alterar o valor de significância do modelo. Logo, o modelo com mais efeitos significativos foi o Impressão. Na tabela abaixo, podemos visualizar os valores de significância:

| Efeito | Vlr Estimado | Erro | P-valor |
|----------------------------|--------------|--------|--------------|
| Intercepto | 927.29 | 129.42 | 2.07e-12 *** |
| Vlr Orçado | -91.04 | 14.50 | 5.84e-10 *** |
| Conta Hardware | -399.46 | 196.49 | 0.-4242* |
| Conta Materiais | -314.73 | 202.17 | 0.11996 |
| Conta Software | -520.77 | 201.16 | 1.39e-05 *** |
| Conta Tel Fixa | -768.37 | 175.61 | 0.60819 |
| Conta Tel Móvel | -109.56 | 213.63 | 0.01299* |
| Conta TI | -782.61 | 314.32 | 0.17479 |
| Conta Internet | -493.82 | 363.56 | 0.07128 |
| Vlr Orçado:Conta Hardware | 40.06 | 22.18 | 0.47439 |
| Vlr Orçado:Conta Software | 18.13 | 25.33 | 0.00991** |
| Vlr Orçado:Conta Tel Fixa | 54.48 | 21.07 | 0.00169 ** |
| Vlr Orçado:Conta Tel Móvel | 66.30 | 21.04 | 0.56429 |
| Vlr Orçado:Conta TI | -15.32 | 26.57 | 0.02330* |
| Vlr Orçado:Conta Internet | 74.37 | 32.72 | 0.18654 |

Tabela 4: Análise do Summary, Fixado Produto Impressão.

O modelo compoem , o valor de intercepto, valor orçado, conta Hardware, conta Software, conta telefone móvel, valor orçado intercepto com conta Software, valor orçado intercepto com conta telefonia fica e valor orçado intercepto com conta TI.

Dando sequência aos testes que devemos fazer esta a análise do resíduos onde podemos observar melhor os dados. E ainda aplicação do teste de normalidade, onde se pode identificar a ausência de normalidade, determinando que o conjunto de dados de uma dada variável aleatória, é modelada por uma distribuição normal ou não.

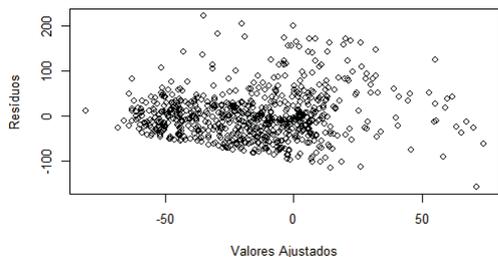


Figura 12: Análise dos Resíduos.

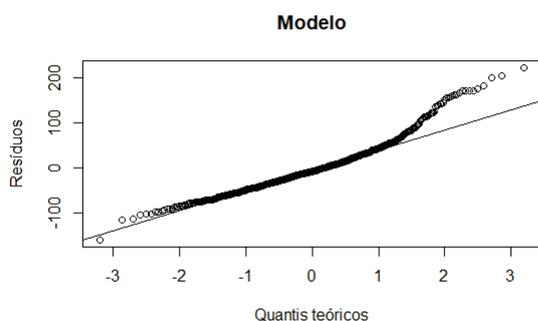


Figura 13: Teste de Normalidade.

Ao realizar o modelo de regressão generalizado, encontramos efeitos não significativos no resultado do summary e também no teste da Anova, que tinha por objetivo comparar a distribuição dos grupos em amostras independentes.

Um dos objetivos da aplicação da ANOVA é realizar o teste estatístico para verificar se há diferença entre distribuição de uma medida entre os grupos, verificando se existe uma diferença significativa entre as médias e se os fatores exercem influência em alguma variável dependente.

Com a variável dicotômica, utilizamos uma segunda metodologia que são os modelos lineares generalizados de efeitos mistos (GLMM). Foi possível testar outros percentuais de diferença entre os valores orçados e realizados, no primeiros momento trabalhamos com o intervalo de 10%

| Efeito | DF | Variância | Residuo | p-valor |
|------------------|----|-----------|---------|---------------|
| Log Orcado | 1 | 42.303 | 510.72 | 7.819e-11 *** |
| Conta | 7 | 11.341 | 499.38 | 0.1244 |
| Log Orcado:Conta | 7 | 6.060 | 493.32 | 0.5328 |

Tabela 5: Análise do Summary, 10%.

| Efeito | DF | Variância | Residuo | p-valor |
|------------------|----|-----------|---------|--------------|
| Log Orcado | 1 | 41.122 | 808.87 | 1.43e-10 *** |
| Conta | 7 | 14.302 | 794.57 | 0.04606 * |
| Log Orcado:Conta | 7 | 4.798 | 789.77 | 0.68466 |

Tabela 6: Análise do Summary, 20%.

Neste tipo de modelagem foi possível observar análise de resíduos e valores ajustados, com uma curva bem coerente para este tipo de dado.

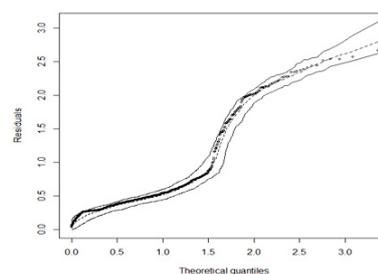


Figura 14: Análise de Resíduos.

Quanto maior o percentual, mais significativo ficam as variáveis, o único cuidado que é necessário ter neste caso, é a análise do negócio. Explicando isso melhor, não é nada interessante para o negócio, demonstrar a probabilidade em 50% de algo acontecer, isso não acrescentaria em nada para respostas ao negócio. Definimos como modelo ideal o que apresenta o % de diferença em 10%.

4. Conclusões e Trabalhos Futuros

Foi possível responder as premissas levantadas no início do trabalho de forma simples e contar com os por métodos estatísticos observando detalhes que estavam intrínsecos dentro dos números.

Os dados gerados foram capazes de validar os gastos em que as áreas orçam e realizam dentro do pacote financeiro que envolve os serviços de TI e Telecom. Essa validação nos fez identificar áreas onde esses números estão de forma erroneamente sendo apresentados.

As análises descritivas nos mostraram que embora os serviços classificados como TI Telecom sejam de base iguais, alguns deles possuem valores já pré estabelecidos e também pudemos identificar qual desses serviços possuem mais gastos.

Estamos falando dos pacote de Software, seguido dos pacotes de Software, Hardware, serviços de impressão, armazenamento e tráfego de dados, serviços de telefonia fixo, serviços gastos com empresas de TI

(consultorias e afins), serviços com telefonia móvel e finalmente materiais de informática, nesta ordem de gasto.

Encontramos diferenças de mais de 12.000%, exemplo esse que fez com púdessemos questionar as equipes responsáveis, sobre a forma como estão escrevendo seu orçamento. Como justificar que queriam gastar por exemplo, R\$ 850,00 em compra de Hardware, e na verdade teve-se um gasto de R\$ 23.450,00.

Todos esses serviços devem ser revistos e analisados, porém o que será feito nesse momento é identificar as prioridades aonde o impacto será mais significativo quando aplicado alguma ação corretiva.

Conseguimos entender a dificuldade das entidades em gerar os valores para o OPEX, classificando as áreas para participarem de treinamentos. O conhecimento profundo da situação pode abrir portas em melhorias na educação e na saúde.

4.1. Trabalhos Futuros

O interessante de se aplicar um trabalho como este é que os números falam por si e o levantamento em questão não precisa ser finalizado em uma única análise. Este trabalho desperta o interesse de aplicar novas técnicas estatísticas afim de buscar novas análises e novos números. Com base no resultado levantado podemos pensar em aplicar técnicas como os Modelos Políticos estes visam analisar itens que apresentam mais de dois escores possíveis (categorias). Basicamente se dividem em duas classes:

Nominais : as categorias não possuem ordenação entre si - itens de múltipla escolha. Ordinais : as categorias possuem ordenação entre si - itens abertos aos quais são atribuídos algum escore [7]

Outra abortagem que poderá futuramente ser investigada seria criar um tratamento diferenciado para dados onde a percentual de diferença entre o valor Orçado versus o valor realizado é muito alto. Entender com cuidado, o motivo dessa variável ter valores muito discrepantes pode nos dar um direcionamento para outro estudo.

Outra extensão de aplicação para o estudo realizado seria a extensão para os dados do CAPEX(despesas ou investimentos com bens materiais)[8] CAPEX (em inglês, CAPital EXpenditure, ou despesas de capital em português) são os fundos que um negócio usa para adquirir bens físicos ou serviços que vão expandir a capacidade da companhia de gerar lucro. Os ativos adquiridos podem ser uma nova posse ou algo que melhore a produtividade de algum equipamento ou

setor. Essas aquisições podem incluir, por exemplo, hardware (como impressoras ou computadores), veículos para transportar bens ou a construção de um novo prédio.

E ainda não podemos destacar o que está em alta no mercado que são os Métodos de Machine Learning para predição desejo de qualquer empresa o poder de prever o futuro utilizando os dados do passado.

5. Agradecimentos

Gostaria de agradecer a Deus por me proporcionar novamente renovar meus conhecimentos, meu esposo e filho por cederem o tempo em família. Aos professores dos Departamentos de Estatística e Computação da UFPR e ao meu professor orientador Cesar Taconeli que pela segunda vez em minha jornada acadêmica dispõem de sua atenção e me proporciona oportunidade de aprendizado.

Referências

- [1] Definição de Orçamento, <https://www.dicio.com.br/>
- [2] Definição de Opex, <https://pt.wikipedia.org/wiki/OPEX>
- [3] Valor Marista, <https://www.grupomarista.org.br/institucional/>
- [4] SEARLE, S.R.(1968) *Another look at Henderson's methods of estimating variance components*. Biometrics, 24, p.749-78
- [5] ZEGER, S.L.;LIANG,K,Y. *An overview of methods for the analysis* Baltimore, v.11 - 1825-1839
- [6] HICKS,C.R. *Fundamental concepts in the design of experiment* 2nd ed. New York, Holt, Rinehart and Winston, 1973. 349p.
- [7] AZEVEDO,Caio <http://www.ime.unicamp.br/>
- [8] Definição de Capex, <https://pt.wikipedia.org/wiki/CAPEX>
- [9] Software R, <http://CRAN.R-project.org>
- [10] Draper N, Smith H. *Applied regression analysis* 3rd Ed. New York: John Wiley and Sons; 1998.
- [11] Bozdongan. H.. *Model selection and Akaike's Information Criterion (AIC)* Psychometrika. v.52, n.3, 345-370, Sep. 1987