



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA
CURSO DE GRADUAÇÃO EM ESTATÍSTICA

DANRLEY DA SILVA ALEXANDRE

RELATÓRIO DE ESTÁGIO SUPERVISIONADO OBRIGATÓRIO II
SERVIÇO SOCIAL DO COMÉRCIO - SESC

FORTALEZA - CEARÁ
2020

DANRLEY DA SILVA ALEXANDRE

RELATÓRIO DE ESTÁGIO SUPERVISIONADO OBRIGATÓRIO II
SERVIÇO SOCIAL DO COMÉRCIO - SESC

Relatório de Estágio Supervisionado Obrigatório II apresentado à Coordenação do Curso de Estatística do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de bacharel em Estatística.

Orientador: Prof. Dr. Gualberto Segundo
Agamez Montalvo

FORTALEZA - CEARÁ
2020

DANRLEY DA SILVA ALEXANDRE

RELATÓRIO DE ESTÁGIO SUPERVISIONADO OBRIGATÓRIO II
SERVIÇO SOCIAL DO COMÉRCIO - SESC

Relatório de Estágio Supervisionado Obrigatório II apresentado à Coordenação do Curso de Estatística do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de bacharel em Estatística.

Orientador: Prof. Dr. Gualberto Segundo Agamez Montalvo

BANCA EXAMINADORA

Prof. Dr. Gualberto Segundo Agamez Montalvo (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. José Roberto Silva dos Santos
Universidade Federal do Ceará (UFC)

Profa. Dra. Jeniffer Johana Duarte Sanchez
Universidade Federal do Ceará (UFC)

AGRADECIMENTOS

À Deus por ter me dado saúde, força e fé. À minha família pelo apoio em todas as etapas da minha vida, em especial meus pais (Sílvia e Amauri), minha avó (Maria de Lourdes) e meus padrinhos (Franck e Eliana), que sempre fizeram o possível para me proporcionar uma boa vida e educação, em um ambiente cheio de amor. Sou muito grato, também, a todos os professores que compõem o Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará, em particular, aos professores Gualberto (Orientador), Maurício e Ana Maria com quem tive o prazer de estudar e que tiveram grande participação em minha formação, tanto acadêmica quanto pessoal. E, por fim, aos meus amigos de graduação, Diego, Lucas Araújo (Maraca), John, Allyson, Eduardo, João Vitor (J.V.), Ramon, Roberto e Vitor (Vitão), pois a convivência, ajuda e apoio deles foi de grande importância nesse período da minha vida.

RESUMO

No relatório, estão presentes as informações relacionadas às atividades desenvolvidas durante o Estágio Supervisionado Obrigatório I e II, na área de Estatística, durante os anos de 2018 e 2019, na instituição do Serviço Social do Comércio - Sesc, além de informações da Empresa e local onde o estágio foi realizado. Os trabalhos foram desenvolvidos acerca dos dados de produção (atividades desenvolvidas pela empresa) e de habilitações, que se referem as matrículas de novos clientes, ou revalidações das mesmas. Especificamente, foi realizado acompanhamento da produção, análise de regressão linear simples para estimação das metas de habilitações e uma análise de pontos influentes na unidade do Centro, pois inicialmente, no Relatório I, foi identificado pontos atípicos que inviabilizavam a aplicação de um modelo de regressão linear simples, para estimar a quantidade de matrículas para o ano de 2020. As atividades desenvolvidas no estágio foram de vital importância para a complementação do ensino e aprendizagem em treinamento prático.

Palavras-chaves: Estágio, Sesc, Regressão Linear Simples.

Sumário

Lista de Figuras	7
Lista de Quadros	8
Lista de Tabelas	9
1 Introdução	10
2 Local de Estágio	11
2.1 Serviço Social do Comércio do Ceará - Sesc-CE	11
3 Objetivos	12
3.1 Objetivos Específicos	12
3.2 Plano de Atividades	12
4 Atividades Desenvolvidas	13
4.1 SGP - Sistema de Gestão do Planejamento	13
4.2 Relatórios de Produção	14
4.3 Acompanhamento de Previsto x Realizado	15
4.4 Prognóstico de Turmas	15
4.5 Análise das Metas de Habilitações	17
4.5.1 Metodologia	17
4.5.2 Ajuste do Modelo	18
4.5.3 Análise de Intervalo de Confiança	23
4.6 Análise na Unidade Centro	24
4.6.1 Análise de Pontos Influentes	25
4.6.2 Ajuste do Modelo	26
4.6.3 Análise de Resíduos	27
5 Considerações Finais	30
Referências	31

Lista de Figuras

1	Prédio do Fecomércio – CE, localização da administração do Sesc-CE. . . .	11
2	Tela inicial de login para acessar o SGP	13
3	Interface do SGP após efetuado <i>login</i>	14
4	Relatório de produção do programa Cultura da unidade do Crato do período de janeiro a maio	14
5	Planilha de previsto e realizado do programa Cultura da unidade do Crato do período de janeiro a maio	15
6	Reta ajustada dos modelos para as unidade do Crato (esquerda) e Iguatu (direita)	20
7	Reta ajustada dos modelos para as unidade de Fortaleza (esquerda) e Centro (direita)	20
8	Envelopes simulados para testar normalidade dos resíduos das unidades do Crato (esquerda) e Iguatu (direita)	21
9	Envelopes simulados para testar normalidade dos resíduos das unidades de Fortaleza (esquerda) e Centro (direita)	21
10	Gráficos de Resíduos x Valores Ajustados, das unidades de Crato (esquerda) e Iguatu (direita)	22
11	Gráficos de Resíduos x Valores Ajustados, das unidades de Fortaleza (esquerda) e Centro (direita)	22
12	Intervalos de confiança para as unidades do Crato (esquerda) e Fortaleza (direita)	24
13	Intervalos de confiança para as unidades do Centro (esquerda) e Iguatu (direita)	24
14	Gráfico para análise de pontos influentes na unidade do Centro	25
15	Reta ajustada do modelo para os novos dados da unidade Centro	27
16	Envelope simulado para teste de normalidade dos resíduos da unidade Centro	27
17	Gráfico de Resíduos x Valores Ajustados para a unidade do Centro	28
18	Intervalo de confiança de 95% da unidade Centro	29

Lista de Quadros

- 1 Prognóstico das turmas PHE do 1º ao 5º Manhã e Tarde em relação à quantidade de alunos inscritos 16
- 2 Prognóstico das turmas PHE do 1º ao 5º Manhã e Tarde em relação à frequência 16
- 3 Prognóstico das turmas PHE do 1º ao 5º Manhã e Tarde em relação à receita gerada 17

Lista de Tabelas

1	Número de habilitações das unidades por ano (em milhares).	17
2	Estimativas obtidas para ajuste do modelo para todas as unidades	19
3	Número de habilitações da unidade Centro após a retirada dos pontos influentes	26
4	Estimativas para o ajuste do modelo para a unidade do Centro	26

1 Introdução

O Estágio Supervisionado é um dos momentos em que o estudante tem para pôr em prática tudo aquilo que aprendeu em sala e investir em sua carreira profissional, enriquecendo o currículo com experiência no mercado de trabalho, o que auxilia na obtenção de um emprego após a formatura. O estágio pode ser desenvolvido em empresas, tanto públicas quanto privadas, instituições de ensino, órgãos públicos, dentre outros. As atividades são desenvolvidas sempre com o cuidado de serem supervisionadas por um profissional da área, no local de estágio, e por um professor orientador, lotado no Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará.

O estágio tem por objetivos principais, a complementação da formação do estagiário, possibilitando o conhecimento prático da profissão, em outras palavras, o estágio é, para muitos, a porta de entrada para o mercado de trabalho, em sua respectiva área.

No presente relatório, serão apresentadas as atividades desenvolvidas no Serviço Social do Comércio - Sesc, durante os anos de 2018 e 2019, sendo elas, acompanhamento da produção (atividades desenvolvidas pela empresa), prognóstico de turmas educacionais, estimação de metas de habilitações, análise dos intervalos de confiança para os valores preditos, das unidades do Crato, Fortaleza, Iguatu e Centro, obtidos por meio de uma análise de regressão simples realizada no Relatório I.

Por fim uma análise mais aprofundada na unidade do Centro, para identificar possíveis pontos influentes e, se possível, remover esses pontos, para tentar ajustar novamente o modelo de regressão linear simples.

2 Local de Estágio

2.1 Serviço Social do Comércio do Ceará - Sesc-CE

Endereço: Rua Pereira Filgueiras, 1070, Aldeota, Fortaleza – CE

- Missão: Contribuir na construção de uma sociedade mais justa, promovendo a melhoria da qualidade de vida do trabalhador do comércio de bens, serviços e turismo, prioritariamente de menor renda, por meio de serviços subsidiados e de excelência (*Sesc,2019*);
- Visão: Ser referência regional na prestação de serviços sociais de excelência, com desenvolvimento da cidadania, sustentabilidade e transferência de tecnologias sociais, até 2020 (*Sesc,2019*);
- Valores: Excelência, ética, qualidade, transparência, integridade, parceria, sustentabilidade, valorização das pessoas, inovação e criatividade (*Sesc,2019*).

Figura 1: Prédio do Fecomércio – CE, localização da administração do Sesc-CE.



Fonte: Sesc, 2019

3 Objetivos

Os objetivos gerais são, a complementação da formação do estagiário, possibilitando o conhecimento prático da profissão, além do desenvolvimento pessoal, possibilitado pelo convívio em um ambiente de integração profissional.

3.1 Objetivos Específicos

- Fornecer suporte ao setor de Planejamento do Sesc-CE para acompanhamento da produção das atividades desenvolvidas por todas as unidades que o compõem;
- Realizar análise estatística de dados, com o fim de auxiliar o acompanhamento da produção;
- Fazer cálculos de projeções da produção para o fim do ano, com a finalidade de tentar estimar se as metas serão alcançadas.

3.2 Plano de Atividades

- Acompanhamento de dados em planilhas de produção;
- Organização e elaboração de apresentações sobre o desempenho das unidades em relação ao previsto x realizado;
- Participação de reuniões avaliativas, junto aos representantes de cada unidade, para debater sobre o desempenho da mesma em relação as metas;
- Construção das metas de matrícula de clientes novos e revalidados para o ano 2020 utilizando técnicas estatísticas (análise de regressão), para estimar os custos necessários para realizar a atividade.

4 Atividades Desenvolvidas

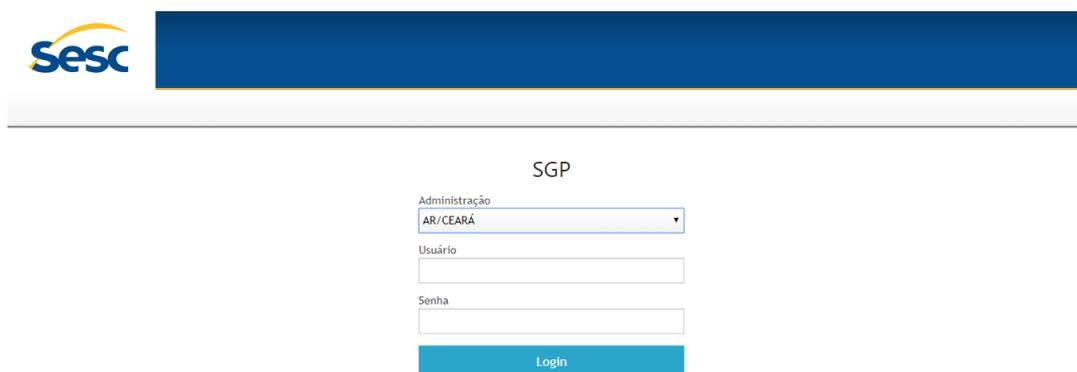
4.1 SGP - Sistema de Gestão do Planejamento

O Sistema de Gestão do Planejamento é um sistema *online* onde todas as unidades que compõem o Sesc lançam a produção, que se refere ao registro das atividades desenvolvidas e ofertadas pela empresa para a comunidade dos trabalhadores do comércio e público geral. Através dele é possível cadastrar projetos, juntamente com suas metas para o ano, produção e orçamento. No começo do ano são definidos os colaboradores responsáveis em realizar o registro dos projetos ao longo do mesmo, depois disso, eles solicitam ao setor de Planejamento que seu perfil seja configurado para que o projeto pelo qual ficou incumbido seja incluído ao seu perfil.

O SGP também gera relatórios de produção, no formato PDF, que é utilizado pelo setor de Planejamento para o acompanhamento mensal da produção de cada unidade.

Na Figura 2, podemos observar a tela inicial de *login* do SGP, nele o colaborador insere seu usuário e senha, fornecidos pela Gerência de Tecnologia da Informação (GTI), para que possa acessar o sistema. Após acessar é apresentada a tela, presente na Figura 3, nela podemos observar uma série de opções, como Usuário, Unidades, Relacionamento com Clientes, entre outros.

Figura 2: Tela inicial de login para acessar o SGP



SGP

Administração
AR/CEARÁ

Usuário

Senha

Login

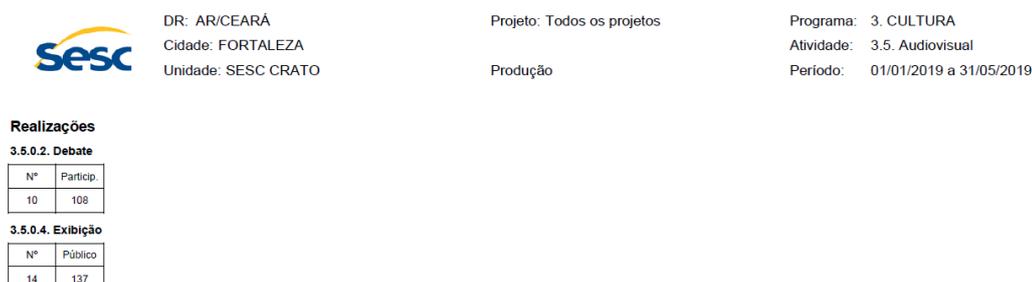
Figura 3: Interface do SGP após efetuado *login*.



4.2 Relatórios de Produção

O SGP gera relatórios de produção apenas em formato PDF, por períodos, como podemos observar na Figura 4. Nesse caso, um relatório de produção da unidade Sesc Crato do programa Cultura e atividade Audiovisual, que consiste em atividades destinadas à produção e difusão de conteúdos audiovisuais, como por exemplo, exibição de filmes ou documentários. Nele os dados são apresentados em forma de quadro, como por exemplo, o quadro Debate, que mostra quantos debates ocorreram no período de janeiro a maio na unidade do Crato, juntamente com a quantidade de participantes que compareceram aos debates, o mesmo pode se aplicar ao quadro Exibição, que apresenta a quantidade de conteúdos exibidos e a quantidade de pessoas que participaram dessas exibições.

Figura 4: Relatório de produção do programa Cultura da unidade do Crato do período de janeiro a maio



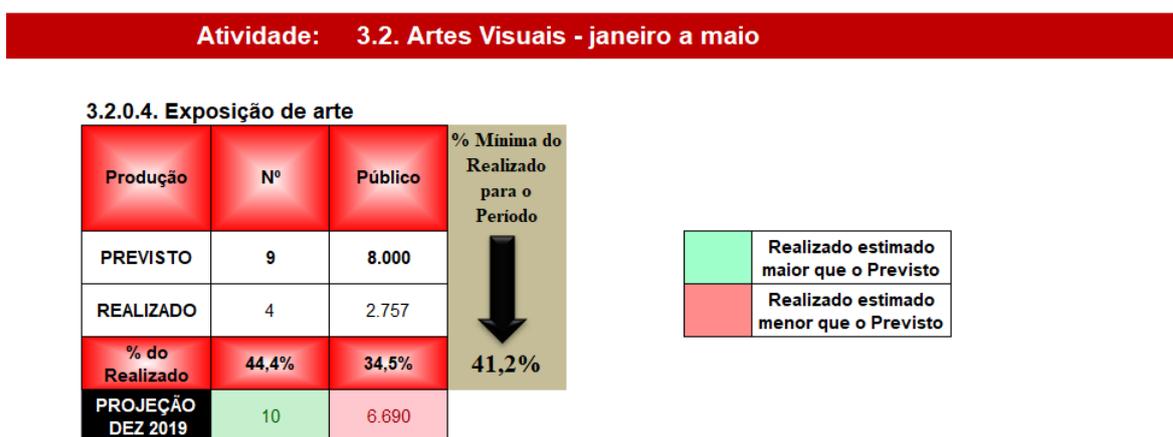
4.3 Acompanhamento de Previsto x Realizado

Um problema encontrado no SGP, é a ausência de um relatório que compare as metas de produção prevista com o realizado no acumulado até cada mês. Para isso foram criadas algumas planilhas em Excel, mais especificamente cinco, uma para cada programa, para que fosse possível realizar esse comparativo, como podemos observar na Figura 5.

Mensalmente são enviadas essas planilhas para o acompanhamento da produção em relação às metas estabelecidas para o ano, para que assim os gestores possam analisar se a unidade por ele gerida está conseguindo atingir as metas ou se é necessário planejar formas de conseguir atingi-las.

Nela podemos observar a Atividade, e comparar o previsto para o ano e a produção até o período. Também temos a porcentagem (%) mínima do realizado para o período, que representa uma estimativa mínima para a produção no referido período em relação a meta, e por fim, uma projeção para dezembro das atividades que serão realizadas, que são calculados levando em consideração a quantidade de dias totais disponíveis e os dias de produção já percorridos.

Figura 5: Planilha de previsto e realizado do programa Cultura da unidade do Crato do período de janeiro a maio



4.4 Prognóstico de Turmas

No período do estágio, em uma ocasião, a gerente do programa Educação da unidade de Fortaleza entrou em contato com o setor de Planejamento, com a intenção de cancelar duas turmas do Projeto Habilidades de Estudo (PHE), são turmas de reforço que tem por objetivo desenvolver habilidades que visam melhorar o desempenho estudantil, alegando

que as turmas estavam com uma quantidade baixa de alunos e que não seria capaz de atingir as metas para o semestre, então foi realizada uma análise exploratória de dados para avaliar a real situação das turmas, no primeiro semestre do ano, e se necessário, realizar o cancelamento das mesmas.

Assim, ao analisarmos Quadro 1, cada turma comporta 25 alunos, entretanto as turmas PHE 1 ao 5 Manhã e PHE 1 ao 5 Tarde estavam com 4 e 8 alunos, respectivamente. A primeira turma estava com um déficit de 21 alunos e 16,0% da capacidade de ocupação da turma, já a segunda turma estava com um déficit de 17 alunos e 32,0% da capacidade de ocupação.

Quadro 1: Prognóstico das turmas PHE do 1º ao 5º Manhã e Tarde em relação à quantidade de alunos inscritos

Prognóstico do 1º Semestre - Alunos				
Turma	Meta de Alunos	Alunos Inscritos	Diferença	% de ocupação
PHE 1º ao 5º Manhã	25	4	-21	16,0%
PHE 1º ao 5º Tarde	25	8	-17	32,0%
TOTAL	50	12	-38	24,0%

No Quadro 2, comparamos a frequência de cada turma, que é contabilizada pela presença do aluno, no semestre, multiplicado pela quantidade de horas em que o mesmo permaneceu na atividade em cada dia, respectivamente. Observamos que ambas as turmas estão com déficit em relação à meta de frequência, sendo que a primeira atingiu apenas 14,6% da meta e a segunda 29,1%.

Quadro 2: Prognóstico das turmas PHE do 1º ao 5º Manhã e Tarde em relação à frequência

Prognóstico do 1º Semestre - Frequência				
Turma	Previsão 2019(Freq.)	Frequência	Diferença	% de Frequência
PHE 1º ao 5º Manhã	9500	1384	-8116	14,6%
PHE 1º ao 5º Tarde	9500	2769	-6731	29,1%
TOTAL	19000	4153	-14847	21,9%

Já no Quadro 3, comparamos a receita gerada por cada turma com sua respectiva meta para o semestre. Temos que a primeira turma atingiu apenas 20,8% da meta prevista e a segunda 41,7%.

Dado que as turmas não atingiram nem 50% das metas definidas para o semestre, foi decidido o cancelamento das mesmas.

Quadro 3: Prognóstico das turmas PHE do 1° ao 5° Manhã e Tarde em relação à receita gerada

Prognóstico do 1° Semestre - Receita				
Turma	Meta de Receita	Receita	Diferença	% de Receita
PHE 1° ao 5° Manhã	R\$ 12.024,00	R\$ 2.505,60	-R\$ 9.518,40	20,8%
PHE 1° ao 5° Tarde	R\$ 12.024,00	R\$ 5.011,20	-R\$ 7.012,80	41,7%
TOTAL	R\$ 24.048,00	R\$ 7.516,80	-R\$ 16.531,20	31,3%

4.5 Análise das Metas de Habilitações

Inicialmente foi ajustado o modelo de regressão linear simples para as metas de habilitações das unidades do Crato, Iguatu, Fortaleza e Centro. O objetivo foi tentar prever quantas habilitações, ou seja, quando o cliente vai à unidade renovar ou realizar um novo cadastro no Sesc, cada unidade realizaria no ano de 2020, afim de ajudar a prever os custos necessários para o próximo ano. A metodologia utilizada foi análise de regressão linear simples.

Tabela 1: Número de habilitações das unidades por ano (em milhares).

Ano	Crato	Fortaleza	Centro	Iguatu
2007	1.997	18.291	12.331	1.095
2008	2.325	24.498	13.901	1.106
2009	2.333	37.760	14.734	1.359
2010	2.540	45.961	16.681	1.561
2011	2.468	53.449	16.351	2.031
2012	2.865	69.375	21.431	2.737
2013	3.528	99.964	11.451	3.008
2014	3.748	110.973	19.914	3.849
2015	3.753	115.826	12.989	4.576
2016	4.014	146.622	20.560	4.711
2017	4.042	138.376	21.126	5.065
2018	4.837	180.997	21.320	5.158

4.5.1 Metodologia

Para este caso foi decidido a utilização da técnica de regressão linear simples, em que X e Y, são as variáveis Ano (explicativa) e Número de habilitações (resposta), respectivamente. Consideramos as variáveis X e Y em n pares, (X_1, Y_1) , (X_2, Y_2) , ..., (X_n, Y_n) . Dado que Y pode ser escrito como função linear de X, pode-se propor um modelo estatístico definido pela equação:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \text{para } i=1,2,\dots,n$$

em que,

- y_i é uma variável aleatória e representa o valor da variável resposta na i -ésima observação;
- x_i representa o valor da variável explicativa na i -ésima observação;
- ϵ_i é uma variável aleatória que representa o erro experimental;
- β_0 e β_1 são parâmetros a serem estimados e que definem a reta de regressão, sendo o intercepto e o coeficiente angular de inclinação da reta, respectivamente;
- n é o tamanho da amostra.

Para que o modelo possa ser utilizado, algumas suposições devem ser atendidas. Sendo elas:

1. A relação matemática entre X e Y deve ser linear;
2. Os valores de x são controlados, ou seja, não é uma variável aleatória;
3. $\epsilon_i \sim N(0, \sigma^2)$, ou seja, segue distribuição Normal com média 0 e variância σ^2 ;
4. Variância constante dos erros, ou seja, homocedasticidade.

Assim, o modelo de regressão linear simples ajustado é dado por:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Para ajustar o modelo foi utilizado o R (*R Core Team, 2018*), que é um software estatístico livre, e especializado em manipulação, análise estatística e visualização gráfica de dados.

4.5.2 Ajuste do Modelo

Para ajustar o modelo de regressão linear simples utilizamos a função $lm()$ do pacote *stats* do *software R*. Após o ajuste, para todas as unidades, obtivemos os resultados

apresentados na Tabela 2, onde podemos observar a estimativa dos parâmetros de β_0 , que representa o intercepto, ou seja, o ponto onde a reta regressora corta o eixo y quando $x = 0$, e β_1 , que representa o coeficiente angular da reta de regressão. Foi fixado um nível de significância (α) de 5%, e podemos perceber que ambos os parâmetros estimados, para as unidades do Crato, Iguatu e Fortaleza, foram muito significativos, pois o valor- p dos dois parâmetros foi bastante inferior a 0,05. A unidade do Centro também foi significativa em relação ao α estabelecido.

Também obtemos o valor de R^2 , que é o coeficiente de determinação. É um valor que nos ajuda a determinar o quão bom foi o ajuste do nosso modelo, juntamente com a análise dos resíduos. Em resumo, R^2 representa a variabilidade dos dados que é explicada pelo modelo de regressão ajustado, em nosso caso, as unidades do Crato, Fortaleza e Iguatu apresentaram estimativas para o coeficiente superiores a 0,90, indicando que para cada uma, o modelo está explicando mais de 90% da variabilidade dos dados, apontando, possivelmente, que para essas unidades o mesmo está bem ajustado. Por outro lado, o R^2 da unidade do Centro foi de 0,3192, ou seja, sugerindo que o modelo está explicando aproximadamente 32% da variabilidade dos dados, indicando que o mesmo não se adequa bem aos dados.

Tabela 2: Estimativas obtidas para ajuste do modelo para todas as unidades

	Parâmetro	Estimativa	Erro Padrão	Valor t	Valor-p
Crato	Intercepto	482.526,08	36.978,99	-13,05	<0,0001
	Ano	241,36	18,37	13,13	<0,0001
	R^2	0,9397	-	-	-
Iguatu	Intercepto	873.668,70	50.466,37	17,31	<0,0001
	Ano	435,62	25,08	17,37	<0,0001
	R^2	0,9647	-	-	-
Fortaleza	Intercepto	28.970.000	1.655.000	-17,50	<0,0001
	Ano	14.440.000	822,60	17,55	<0,0001
	R^2	0,9654	-	-	-
Centro	Intercepto	1.298.508,90	530.089,40	-2,45	0,0343
	Ano	653,60	263,40	2,48	0,0325
	R^2	0,3192	-	-	-

Nas Figuras 6 e 7, podemos observar a reta de regressão estimada, juntamente com os dados, para cada unidade. Ao analisarmos, percebemos que o modelo parece ajustar bem aos dados, nas unidades do Crato, Iguatu e Fortaleza, neles a reta parece bem próxima os dados. Já na unidade do Centro, isso não ocorre, existem muitos pontos afastados da reta.

Figura 6: Reta ajustada dos modelos para as unidade do Crato (esquerda) e Iguatu (direita)

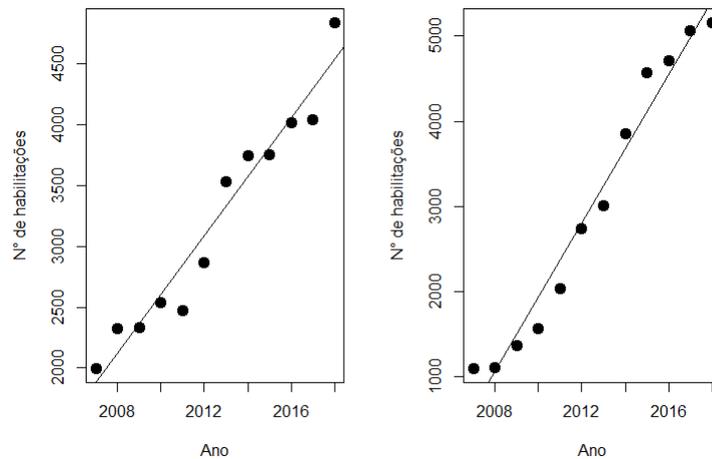
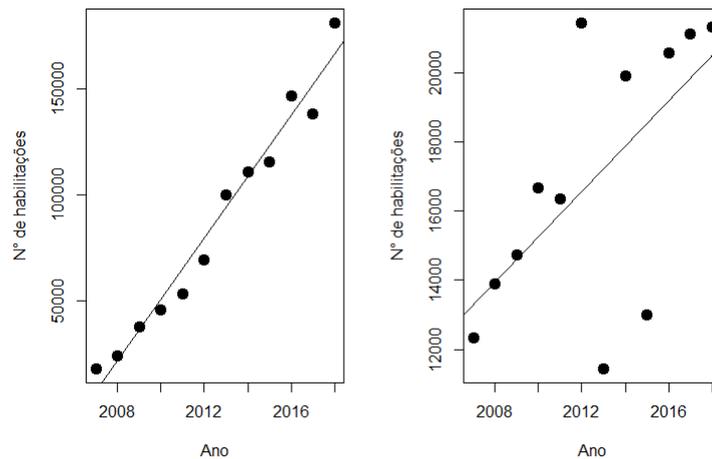


Figura 7: Reta ajustada dos modelos para as unidade de Fortaleza (esquerda) e Centro (direita)



Por sua vez, o R^2 , não é suficiente para determinarmos se o ajuste do modelo foi bom, pois ainda necessitamos testar a suposição de normalidade dos erros e de homocedasticidade de variância, e para isso, foi feito um gráfico de envelope simulado para cada unidade, apresentados nas Figuras 8 e 9, para verificar se os resíduos studentizados seguem uma distribuição Normal, com parâmetros, $\mu = 0$ e $\sigma^2 = 1$. Nas Figuras, podemos

observar que, para as unidades do Crato, Iguatu e Fortaleza, todos os pontos estão dentro do envelope e estão distribuídos em torno da reta. Isso representa um forte indício de que os resíduos studentizados tem distribuição Normal. Entretanto, para a unidade do Centro, percebemos dois pontos fora do envelope e o restante seguindo uma certa tendência, indicando assim que os resíduos não seguem normalidade. Ao levarmos tudo isso em consideração, concluímos que este modelo não foi adequado para a unidade do Centro, sendo necessário uma melhor investigação acerca dos dados, e se necessário aplicação de outros métodos, como por exemplo, analisar pontos influentes.

Figura 8: Envelopes simulados para testar normalidade dos resíduos das unidades do Crato (esquerda) e Iguatu (direita)

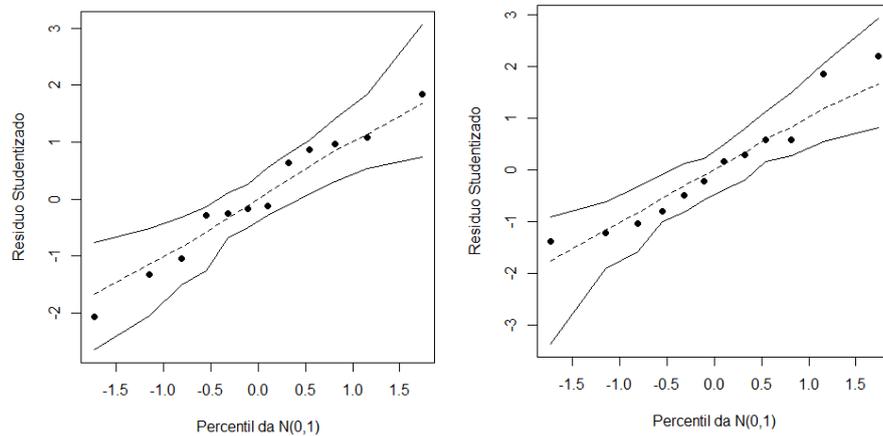
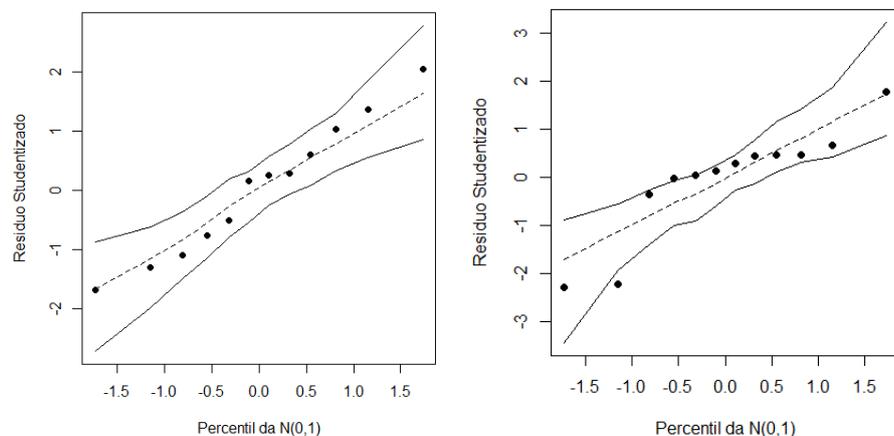


Figura 9: Envelopes simulados para testar normalidade dos resíduos das unidades de Fortaleza (esquerda) e Centro (direita)



Para o verificar a suposição de homocedasticidade foram feitos os gráficos das Figuras 10 e 11, que apresentam os resíduos *versus* valores ajustados, neles podemos observar, que para as unidades de Iguatu, Crato e Fortaleza, os pontos parecem se distribuírem de forma aleatória em torno de $y = 0$, indicando assim que a suposição de homocedasticidade foi atendida. Já na unidade do Centro, os pontos parecem seguir uma tendência, indicando a violação da suposição. Porém, por motivos de estudo, seguimos com as análises.

Figura 10: Gráficos de Resíduos x Valores Ajustados, das unidades de Crato (esquerda) e Iguatu (direita)

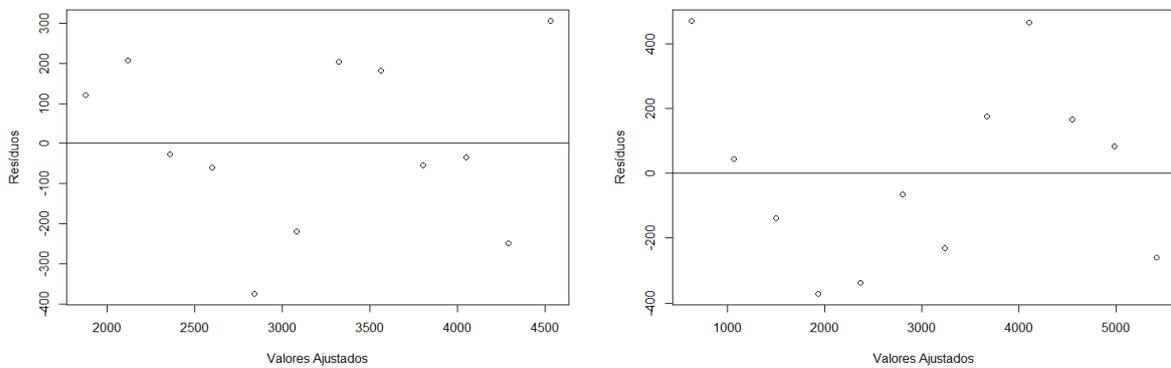
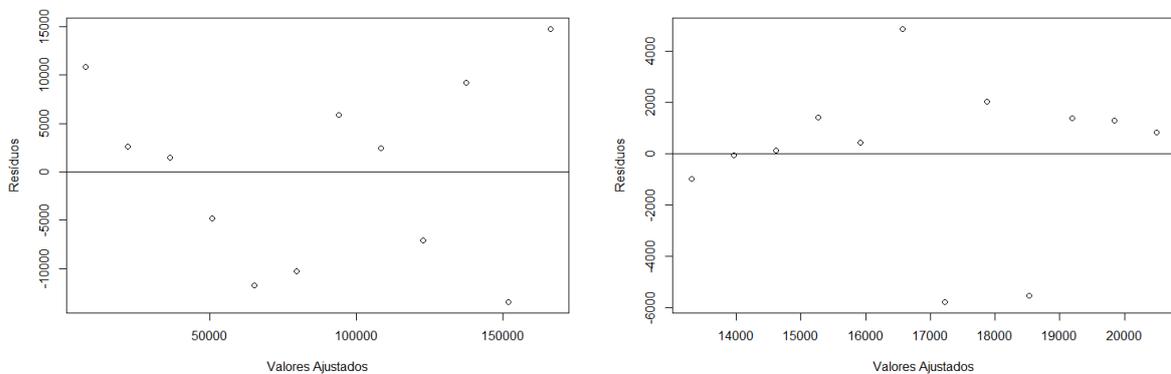


Figura 11: Gráficos de Resíduos x Valores Ajustados, das unidades de Fortaleza (esquerda) e Centro (direita)



Assim, como falado anteriormente, para a unidade do Centro será realizada outra análise. Já para as outras unidades, temos as seguintes estimativas:

- Crato

$$\hat{Y} = -482.526,08 + 241,36 \times Ano$$

$$\hat{Y} = -482.526,08 + 241,36 \times 2020 = 5.021 \text{ habilitações}$$

$$IC(0,95) = [4.676,31 ; 5.352,36]$$

- Iguatu

$$\hat{Y} = -873.668,70 + 435,62 \times Ano$$

$$\hat{Y} = -873.668,70 + 435,62 \times 2020 = 6.284 \text{ habilitações}$$

$$IC(0,95) = [5.827,19 ; 6.749,81]$$

- Fortaleza

$$\hat{Y} = -28.970.000 + 14.440 \times Ano$$

$$\hat{Y} = -28.970.000 + 14.440 \times 2020 = 198.800 \text{ habilitações}$$

$$IC(0,95) = [179.991,50 ; 210.256,30]$$

4.5.3 Análise de Intervalo de Confiança

Analisando os intervalos para as respostas individuais, com 95% confiança, para os valores preditos, nas Figuras 12 e 13, podemos observar, nas unidades do Crato, Fortaleza e Iguatu, que a maioria dos dados estão dentro do envelope ou próximos aos limites superior ou inferior de cada intervalo, também podemos observar que os intervalos estão relativamente estreitos, indicando que os valores estimados pelo modelo dessas unidades estão aparentemente bem precisos e a variabilidade dos dados não é tão grande. Por outro lado, a unidade do Centro apresenta um intervalo de confiança bastante largo, o que já era esperado, pois o modelo não ficou bem ajustado aos dados, como visto anteriormente, implicando em valores preditos não tão precisos.

Figura 12: Intervalos de confiança para as unidades do Crato (esquerda) e Fortaleza (direita)

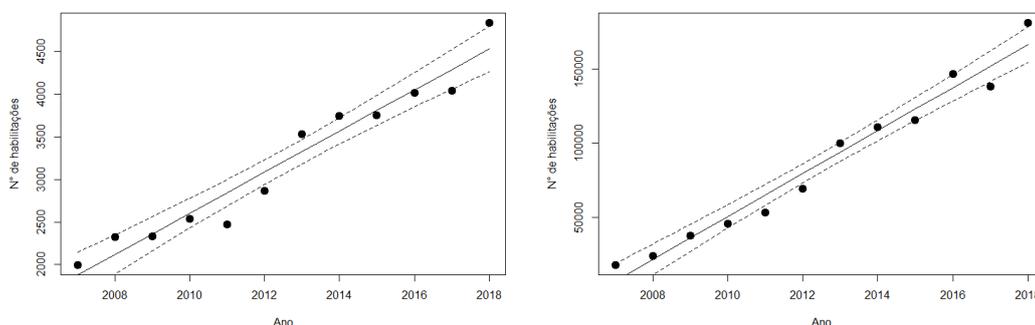
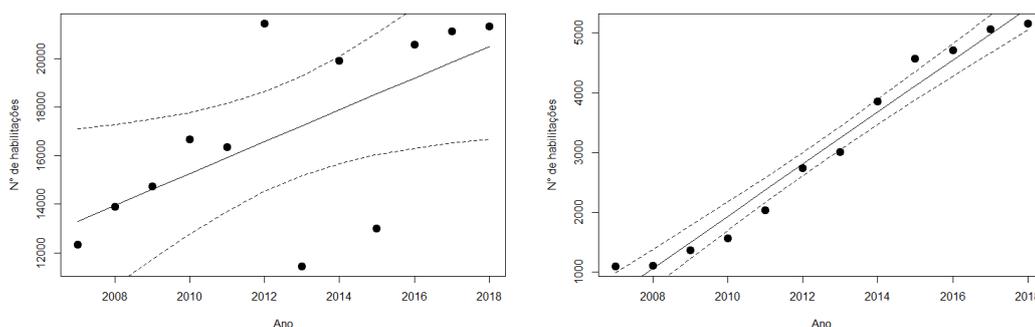


Figura 13: Intervalos de confiança para as unidades do Centro (esquerda) e Iguatu (direita)



4.6 Análise na Unidade Centro

Após uma conversa com o supervisor de estágio para tentar entender o que aconteceu nos anos de 2012, 2013 e 2015, na unidade do Centro, para explicar os valores discrepantes em relação aos demais anos, foi sugerido entrar em contato com a mesma, para obter mais informações acerca dos dados. Na unidade, funcionam uma biblioteca e um restaurante para os clientes. No ano de 2013 o restaurante estava em reforma por um período de 4 meses, então muitos clientes destinaram-se à unidade de Fortaleza para realizar a solicitação ou renovação da matrícula, e posteriormente, utilização do serviço na unidade enquanto a outra estava em reforma.

No ano de 2015, foi a vez da biblioteca ser reformada, por um período de 2 meses, entretanto dessa vez muitos clientes não se dirigiram a unidade de Fortaleza para realizar a matrícula, possivelmente, pelos fatos de que muitos dos clientes que utilizam a biblioteca serem dependentes, ou seja, filhos, esposa, marido, pai ou mãe trabalhadores do comércio,

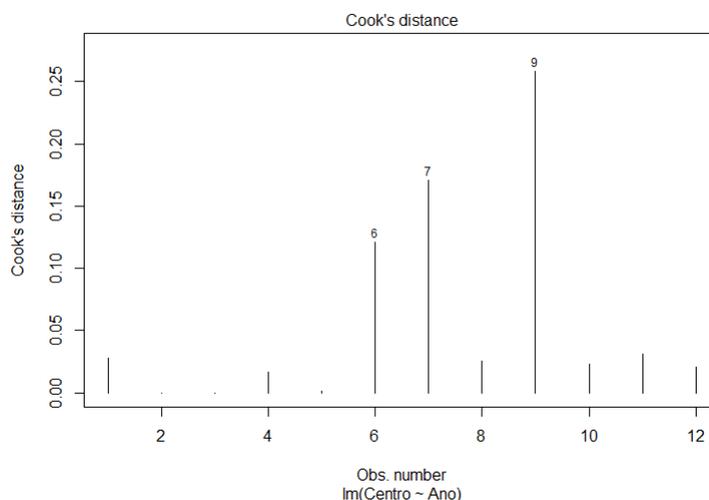
e que também possuem direito de usufruir dos serviços oferecidos pelo Sesc, e o período da reforma foi próximo ao fim do ano, então quando a unidade retomou as atividades no ano seguinte os clientes já foram contabilizados no ano de 2016.

Já no ano de 2012, a unidade de Fortaleza resolveu alocar o setor de Relacionamento com Clientes, responsável pelas habilitações, para outro andar da unidade, processo que durou algumas semanas, nesse período o setor não estava funcionando, então, os clientes se dirigiram a unidade do Centro para realizar o cadastro.

4.6.1 Análise de Pontos Influentes

Tendo conhecimento maior acerca dos dados, foi feita uma análise de pontos influentes, apresentado na Figura 14. Segundo o gráfico de distância de *cook*, podemos perceber que as observações 6, 7 e 9, que representam os dados referentes aos anos de 2012, 2013 e 2015, respectivamente, são provavelmente, pontos influentes.

Figura 14: Gráfico para análise de pontos influentes na unidade do Centro



Após a análise dos pontos influentes e sabendo os motivos dos dados atípicos, foi decidido remover essas observações do conjunto de dados, e novamente ajustar um modelo de regressão linear simples para estimar a quantidade de habilitações na unidade do Centro, para o ano de 2020. Podemos observar o novo conjunto de dados da unidade (Tabela 3) após a retirada dos pontos influentes.

4.6.2 Ajuste do Modelo

Tabela 3: Número de habilitações da unidade Centro após a retirada dos pontos influentes

Ano	Centro
2007	12.331
2008	13.901
2009	14.734
2010	16.681
2011	16.351
2014	19.914
2016	20.560
2017	21.126
2018	21.320

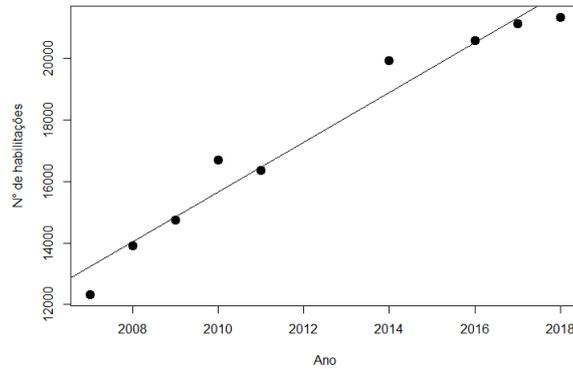
Para ajustar o modelo de regressão linear simples foi utilizado a função $lm()$ presente no software *R* (*R Core Team, 2018*). As estimativas obtidas estão presentes na Tabela 4. Nela podemos observar a estimativa de β_0 (Intercepto) e β_1 (Ano). Foi fixado um nível de significância (α) de 5%, e podemos perceber que ambos os parâmetros estimados foram muito significativos, pois o valor- p dos dois foram muito inferiores a 0,05. Na mesma tabela, podemos observar também o valor estimado de R^2 , que é o coeficiente de determinação do modelo, com ele podemos perceber que o modelo está explicando, aproximadamente 95,5% da variabilidade dos dados, indicando possivelmente que está bem ajustado. Entretanto, para afirmar que o modelo está bem ajustado, ainda é necessário confirmar a suposição de normalidade dos resíduos, entre outras suposições para o modelo.

Tabela 4: Estimativas para o ajuste do modelo para a unidade do Centro

Parâmetro	Estimativa	Erro Padrão	Valor t	Valor- p
Intercepto	-1.609.100	124.552,80	-12,92	<0,0001
Ano	808,30	61,90	13,06	<0,0001
R^2	0,9549	-	-	-

Na Figura 15, podemos observar a reta de regressão estimada, juntamente com os dados da unidade do Centro. Ao analisarmos, podemos perceber que agora o modelo ajusta bem os dados, pois os pontos estão relativamente próximos a reta.

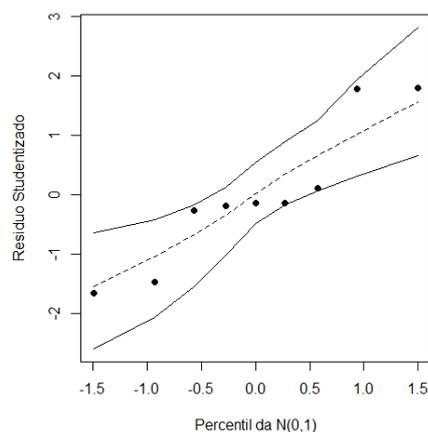
Figura 15: Reta ajustada do modelo para os novos dados da unidade Centro



4.6.3 Análise de Resíduos

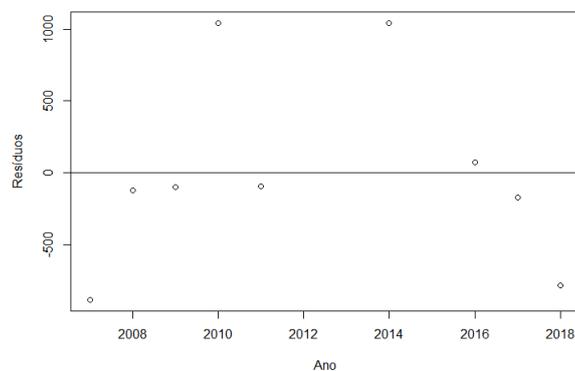
Para testar a suposição de normalidade dos resíduos do modelo, foi feito um gráfico de envelope simulado (Figura 16), para verificar se os resíduos padronizados seguem uma distribuição Normal, com parâmetros, $\mu = 0$ e $\sigma^2 = 1$. Na Figura, podemos observar que todos os pontos estão dentro do envelope e distribuídos em torno da reta, apresentando um forte indício de que os resíduos seguem normalidade. Por fim foi realizado o teste de Shapiro-Wilk (*Shapiro and Wilk, 1965*), que é um teste de normalidade não paramétrico, onde a hipótese nula afirma que a amostra provém de uma população Normal e a hipótese alternativa diz o contrário. Segundo o teste, o valor- p obtido foi de 0,1163, ou seja, não rejeita a hipótese nula, considerando um nível de significância de 5%, indicando que os resíduos seguem uma distribuição Normal.

Figura 16: Envelope simulado para teste de normalidade dos resíduos da unidade Centro



Para verificar a suposição de homocedasticidade, observamos a Figura 17, que mostra o gráfico dos resíduos *versus* valores ajustados, onde podemos perceber que os pontos não parecem se distribuírem de forma aleatória em torno de $y = 0$, violando a suposição. Entretanto, seguimos com as análises, por motivos de estudo, como mencionado anteriormente.

Figura 17: Gráfico de Resíduos x Valores Ajustados para a unidade do Centro



Assim, a equação para o modelo é dada por:

$$\hat{Y} = -1.609.100 + 808,30 \times \text{Ano}$$

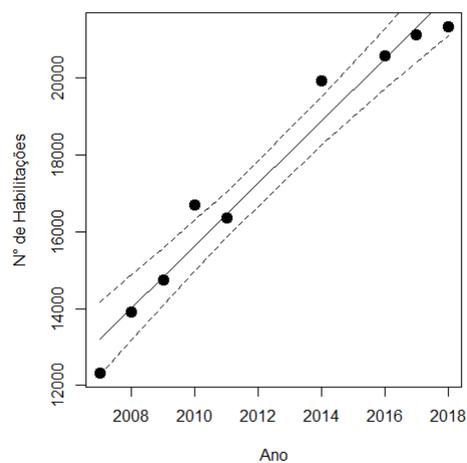
Para o ano de 2020, temos:

$$\hat{Y} = -1.609.100 + 808,30 \times 2020 = 23.666 \text{ habilitações}$$

$$IC(0,95) = [22.450,08 ; 24.994,58]$$

Analisando os intervalos de confiança para os valores preditos (Figura 17), podemos perceber que quase todos os pontos estão dentro do envelope, e o intervalo de confiança está mais estreito, indicando assim que os valores preditos estão relativamente mais precisos, em comparação ao intervalo construído antes de remover os pontos influentes.

Figura 18: Intervalo de confiança de 95% da unidade Centro



Por fim, como a suposição de homocedasticidade foi violada, podemos perceber o que o modelo não é o adequado para ajustar os dados, sendo necessário a aplicação de algum outro método.

5 Considerações Finais

O Sesc é uma empresa que atua como um agente facilitador para a transformação da sociedade, principalmente para a população comerciária, estimulando a melhoria da qualidade de vida, através de ações e serviços oferecidos para a comunidade, nas áreas de educação, cultura, lazer, saúde e assistência.

No presente relatório foram apresentadas algumas técnicas estatísticas de análise descritiva, análise de regressão linear simples e, por fim, análise de pontos influentes, este último utilizado para tentar contornar um problema de ajuste do modelo de regressão linear simples na unidade do Centro, que apresentava algumas observações que não possibilitava a aplicação do mesmo, porém após uma investigação, junto a unidade, descobriu-se o que ocorreu. Então, na tentativa de ajustar o modelo, essas observações foram removidas. Entretanto, como a suposição de homocedasticidade foi violada, percebemos que o modelo não é o adequado para ajustar os dados. Sendo necessária a aplicação de alguma técnica diferente para ajuste.

Por fim, a experiência do estágio contribuiu tanto para o desenvolvimento acadêmico e profissional do estagiário como para o benefício da empresa, apresentando e aplicando técnicas estatísticas para solucionar alguns problemas encontrados no dia a dia.

Referências

- [1] MONTGOMERY, Douglas C.; PECK, Elizabeth A.; VINING, G. Geoffrey. Introduction to linear regression analysis. John Wiley Sons, 2012.
- [2] MORETTIN, Pedro Alberto; BUSSAB, WILTON OLIVEIRA. Estatística básica. Editora Saraiva, 2017.
- [3] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<https://www.R-project.org/>>.
- [4] Sesc – ce. Disponível em : < <https://www.sesc-ce.com.br/>>. Acesso em: 28 de julho de 2019
- [5] SHAPIRO, Samuel Sanford; WILK, Martin B. An analysis of variance test for normality (complete samples). Biometrika, v. 52, n. 3/4, p. 591-611, 1965.
- [6] TOLEDO, Geraldo Luciano; OVALLE, Ivo Izidoro. Estatística básica, Ed. Atlas, 2ª Edição, 1982.