UNIVERSIDADE FEDERAL DO TOCANTINS
CAMPUS DE ARAGUAÍNA
CURSO DE LICENCIATURA EM MATEMÁTICA

MOISÉS DA SILVA SANTOS

ECONOMETRIA: MODELOS DE REGRESSÃO NO ESTUDO DE VARIÁVEIS ECONÔMICAS DE ARAGUAÍNA-TO

ARAGUAÍNA-TO
2021

MOISÉS DA SILVA SANTOS

ECONOMETRIA: MODELOS DE REGRESSÃO NO ESTUDO DE VARIÁVEIS ECONÔMICAS DE ARAGUAÍNA-TO

Monografia apresentada ao Curso de Licenciatura em Matemática, da Universidade Federal do Tocantins - UFT, Campus de Araguaína, como requisito parcial para a obtenção do título de Licenciado em Matemática.
Orientador: Prof. Dr. Rogerio dos Santos Carneiro.

ARAGUAÍNA-TO
2021

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal do Tocantins

S237e Santos, Moisés da Silva.

Econometria: modelos de regressão no estudo de variáveis economica de Araguaína-TO. / Moisés da Silva Santos. – Araguaína, TO, 2021.

59 f.

Monografia Graduação - Universidade Federal do Tocantins – Câmpus Universitário de Araguaína - Curso de Matemática, 2021.

Orientador: Rogerio Dos Santos Carneiro

1. Econometria. 2. Modelos de Regressão. 3. Estudo de variáveis de Araguaína-TO. 4. Estudo econômico do PIB da região. I. Título

CDD 510

MOISÉS DA SILVA SANTOS

ECONOMETRIA: MODELOS DE REGRESSÃO NO ESTUDO DE VARIÁVEIS ECONÔMICAS DE ARAGUAÍNA-TO

Monografia apresentada ao Curso de Licenciatura em Matemática, da Universidade Federal do Tocantins - UFT, Câmpus de Araguaína, como requisito parcial para a obtenção do título de Licenciado em Matemática.

Aprovada em 27 de julho de 2021.

Banca examinadora

Prof. Dr. Rogerio dos Santos Carneiro
Orientador

Prof. Esp. Eloene Sousa Pires Vieira
Examinadora

Kattia Ferreira da Silva
Prof. Ma. Kattia Ferreira da Silva
Examinadora

Sinval de Oliveira
Prof. Dr. Sinval de Oliveira
Examinador

Araguaína / TO
2021

A minha familia.

AGRADECIMENTOS

Agradeço a Deus, por ter me possibilitado esses anos de aprendizagem e me dado foco para concluir.

A minha mãe Liliane, que muito me apoiou e incentivou durante esses anos. Aos familiares e amigos que me incentivaram nessa caminhada tornando-a mais prazerosa.

Aos professores do Colegiado de Matemática que compartilharam seus saberes comigo, me incentivaram e incetivam a buscar mais. Faço este agradecimento no nome do Prof. Dr. Sinval de Oliveira e Profa. Dra. Fernanda Vital de Paula.

Estendo agradecimentos aos professores que colaboraram no período em que estive no PIBID os quais contribuíram de forma significativa na minha formação e experiência, assim faço no nome da Profa. Eloene Sousa Pires Vieira. Aos amigos que encontrei na graduação que muito me ajudaram.

E por fim, ao meu orientador Prof. Dr. Rogerio do Santos Caneiro que com muita paciência, profissionalismo e generosidade me auxiliou e ensinou neste percurso, um excelente professor.

A verdadeira profissão do homem é encontrar seu caminho para si mesmo.

(Hermann Hesse)

RESUMO

Esta monografia visita modelos de regressão para apresentar um estudo Econométrico de valores do PIB no município de Araguaína-TO. O estudo objetivou compreender, por meio da análise de regressão, o grau de relação que existe entre duas ou mais variáveis que fazem a composição principal (agropecuário, prestação de serviços e indústrias) do PIB na cidade de Araguaína-TO, conforme dados organizados pela SEPLAN. Para tal fim, é apresentado os modelos de regressão linear simples e regressão linear múltipla no intuito de encontrar a equação da reta de regressão que melhor explica os valores da amostra, e que por meio desta se torna possível supor previsões ao estudo. Quanto ao método, a pesquisa tem caráter dedutivo e exploratório bibliográfico possibilitando um estudo quali-quanti. Os resultados apontados destacam que alterações feita nesses setores influenciam diretamente no cálculo final deste indicador econômico, se ocorre aumento como consequência da relação direta o PIB cresce e caso ocorra diminuição o PIB também terá decrescimento, afirmando assim, que os modelos de regressão possuem ferramentas de interesse no estudo da Economia, contemplando o objetivo deste estudo.

Palavras-chave: Modelos de regressão. PIB. Análise. Estudo econométrico.

ABSTRACT

This monograph uses regression models to present an Econometric study of GDP values in the municipality of Araguaína-TO. The study aimed to understand, through regression analysis, the degree of relationship that exists between two or more variables that make up the main composition (agriculture, services and industries) of GDP in the city of Araguaína-TO, according to data organized by SEPLAN. For this purpose, the simple linear regression and multiple linear regression models are presented in order to find the equation of the regression line that best explains the sample values, and that through this it becomes possible to assume predictions for the study. As for the method, the research has a deductive and exploratory bibliographic character, enabling a quali-quanti study. The results pointed out that changes made in these sectors directly influence the final calculation of this economic indicator, if there is an increase as a result of the direct relationship, the GDP grows and if there is a decrease, the GDP will also decrease, thus stating that the regression models have tools for interest in the study of economics, contemplating the objective of this study.

Keywords: Regression models. GDP. Analyze. Econometric study.

Lista de Figuras

Lista de Tabelas

Sumário

Capítulo 1

INTRODUÇÃO

A econometria, é o resultado de determinada visão a respeito do papel da economia, esta consiste na aplicação da estatística matemática a dados econômicos para dar suporte empírico aos modelos formulados pela economia matemática e então obter resultados numéricos[1]. Um dos métodos mais convencionais nesta área é a análise de regressão. A regressão, consiste na modelação da relação existente entre duas ou mais variáveis, sendo elas Y variável resposta e X_n variável independente ou preditoras por meio de técnicas estatísticas. De forma genérica, relações funcionais podem ser expressas por:

$$Y = f(X_n)$$

E podemos tomar alguns exemplos, tais como:

- Valor gasto (Y) em função do salário recebido (X);
- Variação do preço de um produto (Y) em função da quantidade de produtos oferecidos (X);
- Variação do salário (Y) em função da taxa de desemprego (X);
- Renda trimestral (Y) em função das despesas (X).

Se efetuarmos um estudo de um banco de dados em que a amostra se verifica em uma quantidade grande, podemos dizer que nem todos estes valores estarão alinhados em uma reta, pois sabemos que muitos valores desse conjunto de dados sofre dispersão por fatores que os perpassam.

Ao nos valermos da idéia de que esses valores mantém uma associação, podemos no direcionar por meio de técnicas estatísticas para a obtenção de uma equação no estudo de regressão, essa equação irá possibilitar a construção gráfica. Então note, que para uma relação funcional dada por $Y = f(X_n)$, teremos valores que se encaixam em x_k $(k = 1, 2, 3, ...n)$ e para cada valor deste um correspondente $Y_k = f(X_k)$ obtendo assim um dado par ordenado de valores que ao plotarmos num gráfico poderemos realizar o estudo da equação de reta visualmente, analisando a relação.

Se levarmos em consideração um relação simples, como:

- Valor gasto (Y) em função do salário recebido (X).

Podemos estudar estes valores por meio da regressão, obteremos uma equação de regressão que pode afirmar a maneira com que estas variáveis estarão relacionadas, e de igual modo quando plotados os valores no gráficos poderá realizar um estudo visual. Quando tratamos de n variáveis preditoras ocorre de existir valores dispersos que podem influenciar no estudo superficial, pois torna-se insociável prever um valor futuro por meio de um padrão se existe valores distantes dos demais, dentro do modelos de regressão esse valor disperso o qual chamamos de erro (e_i) não terá tanta influência no estudo das variáveis econômicas do município de Araguaína.

O território de Araguaína é uma Microregião do estados do Tocantins, com uma população estimada segundo o censo do Instituto Brasileiro de Geografia e Estatística (IBGE) 2020, de 183381 pessoas, e área territorial de 4.004,646 km^2 sendo o segundo mais importante centro econômico do Estado, atrás apenas da capital.

A inquirição por regressão é usualmente tido por convencias no modo de estudo, para o viés deste trabalho a regressão terá grande valor na leitura visual (digrama de dispersão), nas técnicas para encontrar a equação da reta, no estudo gráfico-visual da equação plotada e por meio dela uma análise com intuito de identificar o modelo à fazer predições.

Demasiados modelos de regressão são constantemente usados em estudos de variáveis parcimoniosa, afim de representar como estas se relacionam por meio de uma equação definida por essa associação. A situação norteadora desta pesquisa é o estudo do PIB (Produto Interno Bruto) Araguainense. O PIB é a soma de todos os bens e serviços finais que são produzidos geralmente no período de um ano por um país, estado ou cidade o qual é calculado na moeda nacional[14]. Esse indicador econômico foi criado para entender o impacto populacional sobre a produtividade e atualmente tem como finalidade fazer um diagnóstico local para evidenciar possíveis falhas no crescimento econômico.

Nesse sentido indagamos, que estudos de regressão podem ser feitos com dados do PIB de Araguaína-TO para servir como auxilio a qualquer planejamento econômico da cidade? Para um melhor delinear no decorrer da pesquisa, foi necessário investigar as seguintes questões secundárias: Quais setores do PIB pode ser fomentados para se obter um aumento? Quais variáveis mais auxiliam no PIB de Araguaína-TO? Se uma ou mais dessas fontes do PIB não produzir, o que poderá ocorrer posteriormente? Ou pelo contrário, se aumentarem o valor anual do PIB na cidade, o que irá ocorrer?

São perguntas que nortearam o destrichar do trabalho através do estudo dos modelos de regressão linear simples e linear múltipla. O qual objetivou compreender, por meio da análise de regressão, o grau de relação que existe entre duas ou mais variáveis que fazem a composição principal (agropecuário, prestação de serviços e indústrias) do PIB na cidade de Araguaína-TO. Sabendo também que alguns destes modelos são próprios para fazer uma predição (prever valor futuro).

Assim, destacamos alguns objetivos essenciais para o êxito ao fim do trabalho. Inicialmente caracterizar alguns conceitos fundamentais para um estudo ecométrico, por meio de análises de regressão os quais subsidiaram o foco deste trabalho, para que obtivéssemos ferramentas para, desenvolver um estudo teórico sobre regressão linear simples e múltipla a fim de compreender algumas técnicas a serem usadas no exame, e posteriormente aplicar esses conceitos de regressão nas variáveis reias da amostra para compreender por meio de uma análise a realidade do PIB do município.

Compreendemos que dentro das variadas situações econômicas temos fatores/variáveis que podem estar associadas, sendo assim podemos explicar algum acontecimento do ramo ou fazer alguma previsão futura por meio da regressão. Com essa idéia conseguimos direcionar o estudo sobre o PIB no consideramos os setores de fomento.

Vimos no decorrer do segundo capítulo dessa pesquisa os pontos históricos mais importantes sobre as ciências que usaremos aqui, no capítulo seguinte relembramos de maneira breve conceitos da ciência Estatística que dão subsídios ao estudo principal. No quarto capítulo estudamos os modelos de regressão linear simples e múltipla, e concretizamos nosso objetivos, no quinto capítulo onde conseguimos analisar e entender como esses setores mantém relação com o PIB, onde de fato vimos que há um grau de relação significativo onde as alterações dos setores acarreta alterações no PIB final.

Capítulo 2

NOTA HISTÓRICA

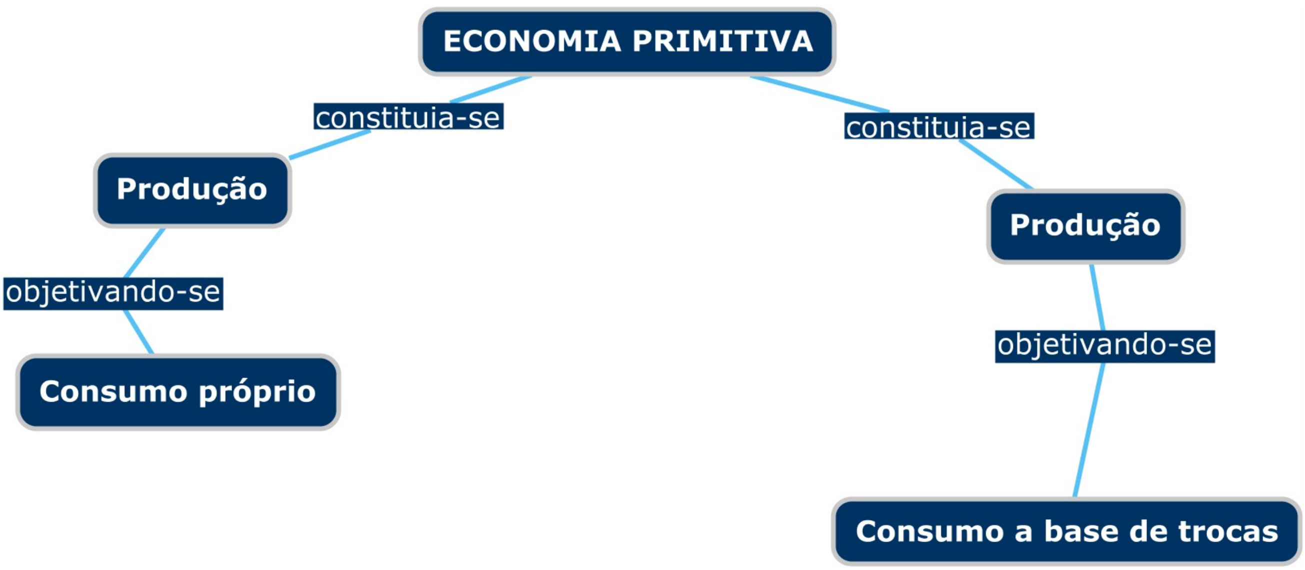
Neste capítulo, iremos discutir um breve resumo do contexto histórico do surgimento de importantes temas neste trabalho. Deste modo, as seções abaixo serão destinadas aos relatos para fundamentar e dar embasamento a este trabalho.

2.1 Uma Breve História da Economia

Observando a história em seus termos primitivos, podemos entender que os fatos econômicos sempre existiram, o que não existiu foram os fatos formais e organizados atualmente[2]. Com o objetivo de conquistas cotidianas o homem passou a sistematizar e organizar os termos econômicos. A economia primitiva "essencialmente se baseava na agricultura, tinha um caráter bastante heterogêneo ou seja, não havia uma unidade de organização economica"[3](p. 6), este momento econômico se baseava "no controle da produção de bens de subsistência"[3](p. 6).

A Figura 01 identifica os momentos da economia primitiva, onde nota-se a intensa atividade de produção pelas sociedades da época com finalidade o consumo próprio, e dando início a produção para trocas, onde a economia em si começa ter mais significado.

Figura 01 - Economia Primitiva



Fonte: Construção própria.

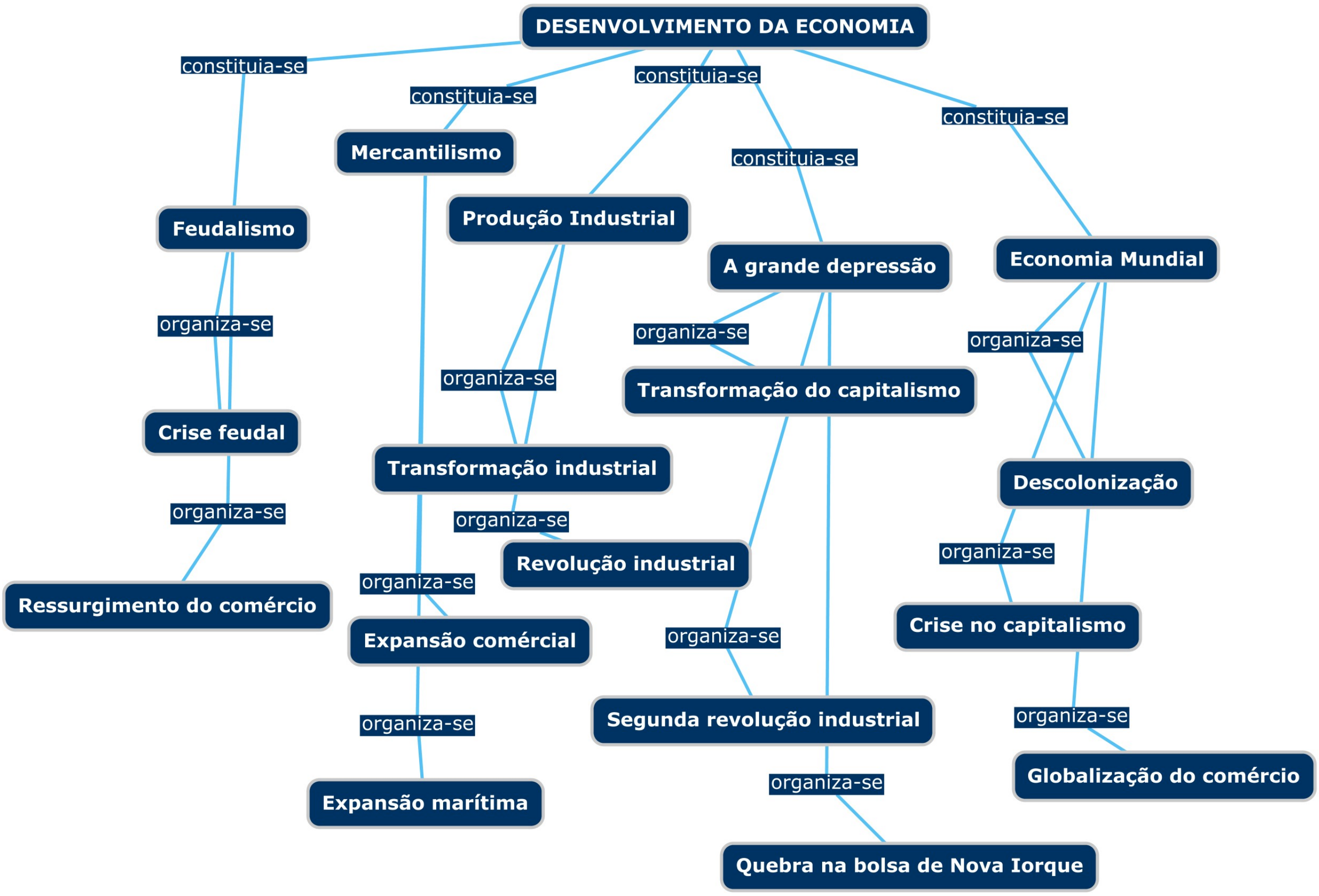
Um momento datado importante na história da economia é o feudalismo, é usual identificar-se o feudalismo como o que chamamos de Idade Média[8]. Nesse momento a economia sofre uma mudança nas atividades comerciais, onde a moeda perde espaço e a ruralização entra em cena com a crise que ocorre no Império Romano.

Com o surgimento do comércio na Europa Ocidental, o feudalismo sofre impacto e então uma crise chega ao sistema feudal. Posteriormente, ocorre a chamada expansão marítima entre os séculos XIV e XV que visava o lucro como foco, e uma expansão comercial dada pelo renascimento urbano.

Nesse intervalo acontecem transformações na produção industrial, fruto do mercantilismo e expansão comercial Europeia. A revolução industrial foi um período de grande desenvolvimento no século XVIII com início na Inglaterra. O surgimento das indústrias ocasionou mudanças na economia mundial que engajou a produção de mercadorias e acelerou a exploração de recursos naturais bem como movimentou transformações na oferta de empregos.

Esse marco, é um passo importante no manuseio e desenvolvimento tecnológico, abrindo espaço para mão de obra e também ao lucro. Desde então, com as crises e guerras que ocorreram e as mudanças nos modos de governos mundial, a economia se modificou e vem crescendo a passos largos.

Figura 02 - Desenvolvimento da Economia



Fonte: Construção própria.

Como visto nessa breve construção histórica, os passos para o desenvolvimento desta importante ciência, se deu na forma como a sociedade sentia necessidade. Dentro disso, uma definição que pode-se assumir à economia é como sendo a ciência que visa o estudo de como homem decide empregar seus recursos. A partir do século XVIII a economia passou a crescer de forma mais ordenada buscando de forma racional o manuseio dos recursos da sociedade, e podemos notar no decorrer desse processo que o surgimento da moeda foi uma consequência natural no desenvolvimento do comércio, pois havia dificuldades nas ofertas de trocas e então o dinheiro “instrumento da circulação das mercadorias”[5](p. 5) entra em cena como forma de pagamento, e passa por processos até chegar aos diferentes tipos de moedas e seus valores como temos atualmente.

2.2 Uma Breve História da Estatística

Entendemos como já mencionado anteriormente, a Estatística como uma ciência independente emergida da grande área matemática repleta de ferramentas e técnicas de estudos. Pode-se até pensar que as técnicas estatísticas nasceram neste mundo contemporâneo, em que se valoriza cada vez mais a rapidez e a agilidade das informações, de um mundo onde o avanço tecnológico (através da criação de computadores que processam uma imensa quantidade de dados em um piscar de olhos) é constante. Entretanto, a utilização da Estatística no auxílio para a tomada de decisões é averiguada também no mundo antigo, e há indícios do seu uso são encontrados até na Era antes de Cristo.

Muitos anos antes de Cristo começaram a surgir às necessidades do conhecimento numérico, pois contar e recensear sempre foram uma preocupação em todas as culturas. O primeiro dado estatístico disponível foi o de registros egípcios, de presos da guerra na data de 5000 a.C., à 3000 a.C. No ano de 2238 a.C. o Imperador da China Yao, ordenou que fosse feito o primeiro recenseamento com fins agrícolas e comerciais. Em 600 a.C. no Egito todos os indivíduos tinham que declarar todos os anos ao governo de sua província a sua profissão e suas fontes de rendimento, caso não a fizessem seria declarada a pena de morte[6].

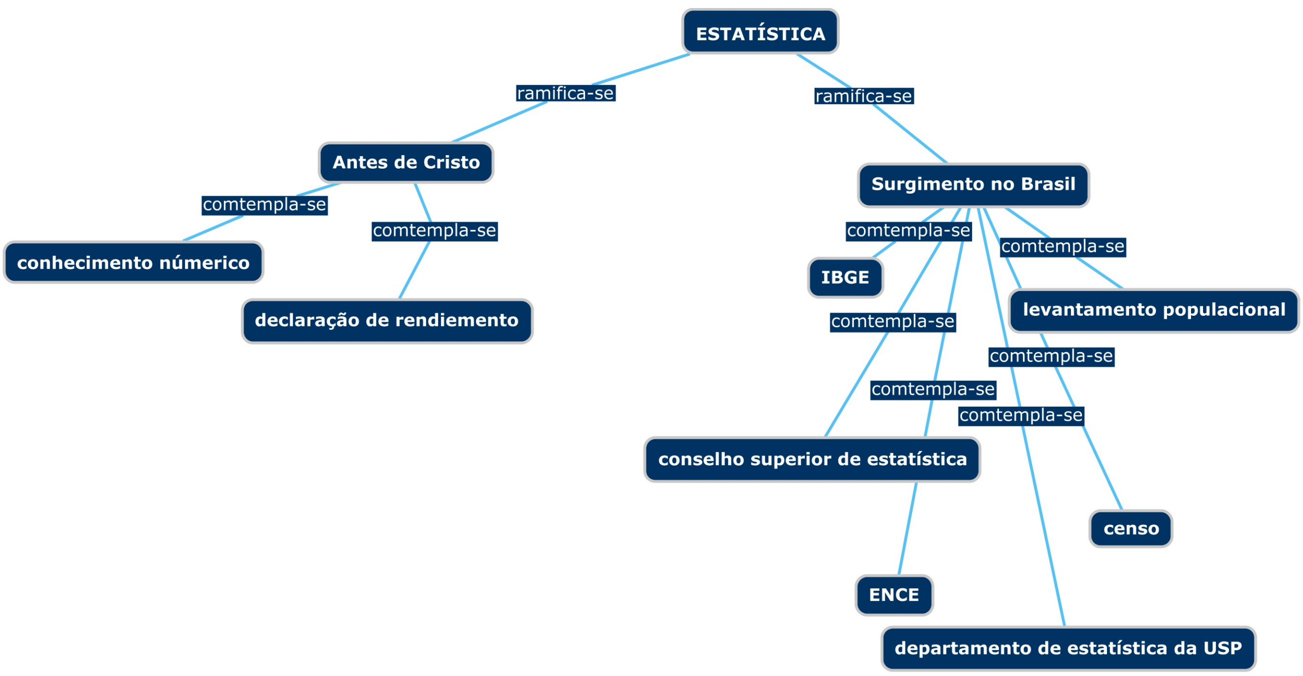
No Brasil, a Estatística tem sua história ligada à história do Instituto Brasileiro de Geografia e Estatística – IBGE, de acordo com o calendário comemorativo dos 50 anos de sua fundação, quem primeiro coordenou e sistematizou atividades ligadas a levantamentos censitários, foi a Diretoria Geral de Estatística, criada em agosto de 1872, data do "primeiro Recenseamento Geral do Império do Brasil". No período anterior a esta data (1750 - 1872), a Coroa Portuguesa era quem determinava levantamentos populacionais, realizados precariamente, com o objetivo maior de "conhecer a população livre e adulta apta a ser usada na defesa do território". Ainda de acordo com o Calendário, foi criado, em 1907, o Conselho Superior de Estatística, com vistas a padronização de conceitos e apuração de resultados em todo o território nacional.

Em 1934, foi criado o Instituto Nacional de Estatística, que só passou a existir de fato em 1936, mudando em 1938 para Instituto Brasileiro de Geografia e Estatística - IBGE, quando os serviços geográficos foram a ele vinculados. Foi a partir de 1940 que se iniciaram os "modernos censos"decenais, não ocorrendo apenas o de 1990 (foi adiado para 1991), devido à "falta de recursos"alegada pelo Governo Collor, antes disso ocorreram os de 1872, 1890, 1900 e 1920. Hoje ele é chamado de Fundação Instituto Brasileiro de Geografia e Estatística - IBGE, sendo integrante da Administração Federal, subordinado diretamente à Secretaria de Planejamento e Coordenação Geral da Presidência da República - SEPLAN/PR, tendo seu Estatuto aprovado pelo Decreto número 97.434 de 05 de janeiro de 1989. As finalidades básicas são a pesquisa, a produção, a análise bem como difusão de informações e estudos de natureza estatística, socioeconômica, na visualização de recursos e condições de recursos naturais importantes ao conhecimento da realidade econômica e social[7].

O primeiro curso de Inferência dado no Brasil ocorreu em 1947, muito embora somente em 1953 duas Escolas iniciassem o ensino de Estatística no Brasil: uma, a Escola Nacional de Ciências Estatísticas - ENCE, criada pelo IBGE nesse mesmo ano, com vistas a contribuir no cumprimento de sua missão institucional. A outra, também fundada em 1953 e mantida pela Fundação Visconde de Cairú, era a Escola de Estatística da Bahia. Em 1948 ocorreu a 1ª mesa redonda sobre o ensino de estatística e a partir desta data houve um crescimento no interesse deste assunto em várias comunidades científicas no mundo todo. A educação estatística surgiu da adaptação às propostas da UNESCO que, moveu incentivos ao desenvolvimento de pesquisas sobre as necessidades para a educação estatística, assim como a formação de um programa internacional para vir ao encontro destas necessidades[9]. Com este propósito foram criados comitês e associações com o objetivo de promover e fomentar estudos e debates sobre a educação estatística. Como resultado deste movimento, surgiu em meados dos anos 70 o ISI (Instituto Internacional de Estatística), criado com o objetivo ampliar e incentivar as pesquisas na área de educação estatística. Em 1970 surgiu a ideia de acrescentar a Estatística no ensino da matemática nas escolas, na primeira conferência do Comprehensive School Mathematics Program, onde foi proposto que no currículo da matemática fosse inclusa noções de estatística e probabilidade desde o curso secundário.

Em 1961, Jerzy Neyman[1] permanece por um mês em São Paulo, onde propôs a criação de um Departamento de Estatística na Universidade de São Paulo - USP. Que foi concretizada onze anos depois, em 1972, com a criação do Departamento de Estatística e o Curso de Bacharelado em Estatística, formando sua primeira turma em 1975.

Figura 03 - Desenvolvimento da Estatística



Fonte: Construção própria.

[1]Nascido em 18 de abril de 1894 na Rússia, matemático e estatístico, auxiliou no estabelecimento de teorias estatística para teste hipóteses, e é tido como principal fundador da estatística moderna.

Com esse sucinto levantamento histórico da constituição da estatística, constata-se que inicialmente ela não era reconhecida como ciência porém, era comumente usada nos atrelados sociais e políticos-econômicos. Com passar do tempo ela foi ganhando força pelos estudiosos e pelo seu papel na sociedade, e grandes momentos datados surgiram como a criação do IBGE e sua importante função nos meios econômicos em geral. A estatística atualmente é um ciência independente, associada à vários tipos de estudos qualitativos e quantitativos, a qual tem crescido e se desenvolvido notoriamente.

2.3 Uma Breve História da Econometria

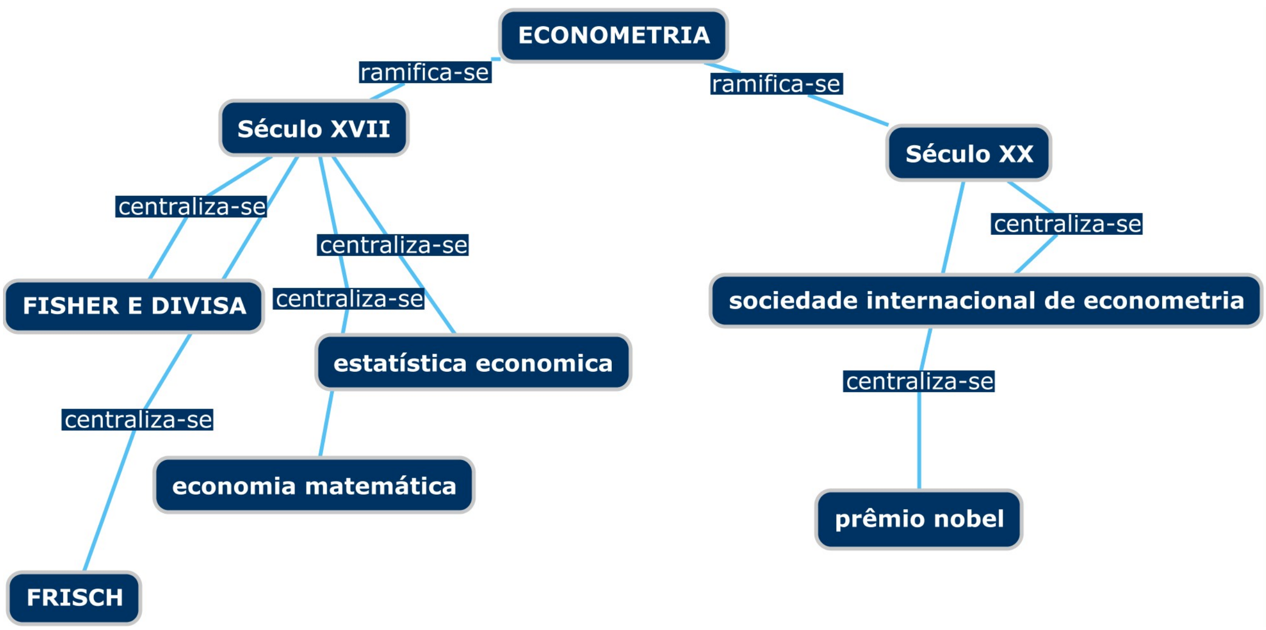
Sabemos que a relação entre matemática e economia se iniciou a bastante tempo, "Como se poderá ver, a econometria não é coisa de agora, mas do século XVII, que já iniciava todo processo, quanto à utilização da matemática e estatística dentro dos trabalhos filosóficos"[2](p. 18). Os primitivos lutavam pela sobrevivência, daí, inicia-se a busca por um arranjo organizacional com finalidade em encontrar o bem-estar.

Como já visto da primeira parte deste capítulo a economia começa se desenvolver de forma estruturada partindo das necessidades da sociedade. Lado a lado da economia caminhavam outras ciências tais como a matemática e estatística. Assim como é hoje, a matemática era importante na tomada de decisões precisas e a estatística por sua vez conseguia explicar a realidade por seu poder explicativo. Nesse caminho surge a economia matemática com objetivo de representar por meio de equações "relações que possam existir entre variáveis representativas da vida econômica"[2](p. 20). E a estatística econômica que apontava o grau de precisão que existia nas afirmações.

Foi nesse sentido, e no atrelar da economia matemática e estatística econômica que surgiu a econometria no século XVII. Considera-se pioneiros da econometria Irving Fisher[2] e Françóis Divisa. Posteriormente com muitas evoluções da econometria ocorre a criação da Sociedade Internacional de Econometria em 1930, e em seguida no ano de 1960, Ragnar Frisch introdutor do termo econometria e Jan Tinberg, ganham o prêmio nobel de econometria.

[2]Nascido em 27 de fevereiro de 1867 no Estados Unidos, foi economista com grandes contribuições à economia tais como a equação de fisher.

Figura 04 - Desenvolvimento da Econometria



Fonte: Construção própria.

Dentro dessa introdução histórica vista na Figura 04 a economia se ligou a matemática originando a econometria em que se visualiza sua associação com a ciência matemática e a ciência estatística a fim de quantificar e explicar os dados e a forma como estão relacionados e segundo a sua definição (definição de econometria), temos que o método mais usual para econometria é o positivista "pois, a estruturação matemática, com a estimação dos parâmetros das equações formalizadas, pode-se tirar conclusões sobre a realidade econômica"[2](p. 20), assim entendemos que modelos algébricos no formato de equações auxiliam em diversos estudos.

2.4 Uma Breve História da Regressão

Uma das principais ferramentas é o estudo de regressão que serve como grande auxílio no estudo da economia por viés estatístico. Em 1854 com a publicação de Darwin sobre A origem das espécies, Francis Galton (1822-1911) propõe a seleção artificial e cria Eugenia.

Tinha ele o objetivo de melhoramento da espécie humana a partir de casais selecionados. Eugenia[3], além de estimular a procriação de casais por meio da matemática, estatística e biologia, "previa evitar a reprodução de casais com características degenerativas"[10](p. 29). Galton também entendia que as crianças herdavam características intelectuais de seus antecedentes.

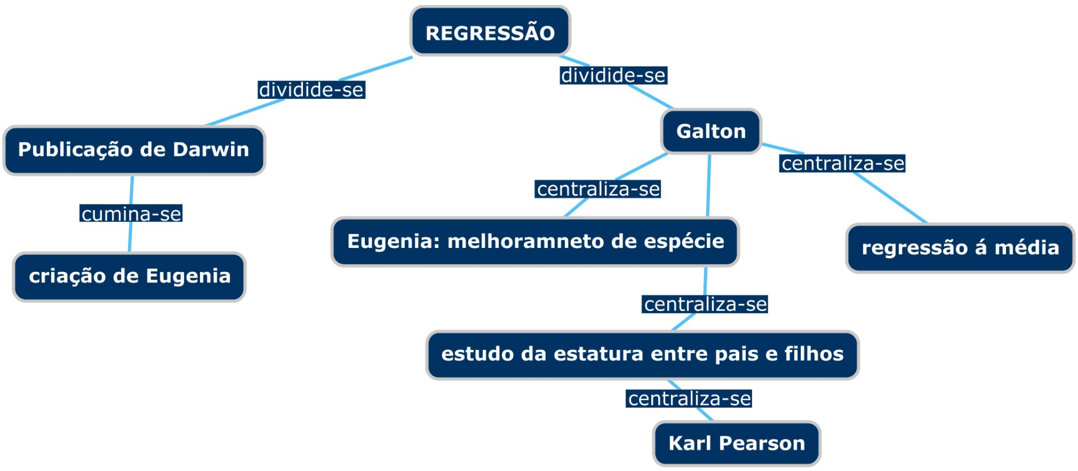
Com 6 anos de estudos e 900 registros familiares, Galton fascinado por medidas descobre a regressão à média. Nesse sentido ele explica o "controle da estatura entre pais e filhos"[10](p.

[3] Teória que busca produzir uma seleção baseada na genética.

29). Se a regressão à média não acontece, notamos que os filhos de pais altos seriam mais altos; os netos mais altos ainda e assim os indivíduos mais altos seriam cada vez mais altos[11]. Com esse estudo Francis Galton mostrou que a altura dos filhos não tinha influência sobre a altura dos pais, mais tinha influência sobre a média da altura da população.

A lei de regressão universal de Galton é então confirmada por Karl Pearson que ampliou e intensificou o mesmo estudo. Atualmente, entende-se a análise de regressão como um vasto campo de técnicas estatísticas com finalidade de modelar relações existente entre variáveis quantificadas e fazer predições através dessa modelação.

Figura 05 - Desenvolvimento da Regressão



Fonte: Construção própria.

O modelo de regressão determinado por Galton foi modernizado até chegar o que conhecemos atualmente, e seus estudos como sintetizado na Figura 05 foram intensificados por Karl.

Como visto no decorrer deste capítulo, a econometria a qual se constitui como foco deste estudo pelas ferramentas matemáticas e estatísticas se contextualizam historicamente em momentos como mencionados anteriormente. Este atrelar, se sistematizou e se intensificou podendo nortear estudos em diversas áreas a despeito da economia, possibilitando previsões e uma leitura conclusiva sobre a realidade.

Capítulo 3

NOÇÕES PRELIMINARES

Para atrelar duas ciências importantes no estudo de variáveis até então desconhecidas, usar-se-a os seguintes modelos de regressão: regressão linear simples e regressão linear múltipla. Para isso, alguns conceitos e definições serão vistos e lembrados abaixo, para dar base a toda fundamentação teórica seguinte.

3.1 Conceitos Fundamentais e Definições

Algumas definições serão fundamentadas antes de adentrarmos ao tema principal. Os conceitos serão apresentados conforme[12] e[13].

Definição 3.1. Estatística. *Ciência que trata da coleta, organização, análise e interpretação dos dados para a tomada de decisões.*

A estatística ciência em foco neste trabalho, dispõe-se de um alto valor explicativo. Como visto na secção de sua história a coleta, organização e análise de dados sempre foi muito usada em ações políticas e economicas, as mesma compõem ferramentas essenciais da estatística.

Dentro da Estatística algumas ferramentas de estudo devem ser ainda definidas:

Definição 3.2. População. *População é o conjunto de todos os elementos dos quais se obtém as informações.*

Olhando para 5.8 nossa população séria todas composições do PIB colocadas, ou seja, seria o 'todo' dessa tabela.

Definição 3.3. Amostra. *É um subconjunto da população.*

As partes direcionadas às colunas podemos considerar amostra como em 3.3, sendo assim uma amostra da produção do PIB Araguainense, é a produção agropecuária (subconjunto).

Definição 3.4. Variável. *Característica à qual se refere a informação.*

Quando mencionamos características aos valores da amostra por exemplo, estamos colocando-os no papel de variável. De forma geral, todo o conjunto de dados é chamado de população, o subconjunto (interesse de estudo) chama-se amostra e as característica que remetem aos dados analisados (amostra) é a variável.

3.2 Variável Aleatória

Quando estudamos a descrição de dados, entendemos que os meios disponíveis para análise das variáveis quantitativas são mais ricos[12], dessa forma, para termos um melhor resumo dessas variáveis aleatória discretas, iremos nos valer de algumas ferramentas que terão importância nos estudos de regressão.

3.2.1 Esperança Matemática

Sabemos por definição que X é variável aleatória discreta[12]. Assim, assumindo os valores $x_1, x_2, ..., x_n$ como valores da amostra o valor médio ou esperança matemática de X[12] é dado por:

$$E(x) = \sum .X_i.P(X_i) \tag{3.1}$$

Assim definimos algumas propriedades considerando X e Y variáveis aleatórias e Z constante:

- $E(Z) = Z$
- $E(X + Z) = E(X) + Z$
- $E(ZX) = ZE(X)$
- $E(X + Y) = E(X) + E(Y)$
- $E(X.Y) = E(X).E(Y)$

São estas propriedade relevantes para o manuseio no trabalho de variáveis aleatórias no status condicional.

3.2.2 Variância e Covariância

Temos que a variância é uma medida de análise para dispersão da distribuição, ela se associa a variação do conjunto de dados, ou seja, o quão distantes estão em relação a média. Já a covariância, é a determinação em medida do grau de interdependência das variáveis aleatórias.

Calculamos a variância por:

$$V(X) = E[X - E(X)]^2 \tag{3.2}$$

e determinamos a covariância pela fórmula:

$$Cov(X, Y) = E[X - E(X)][Y - E(Y)] \tag{3.3}$$

Quando tratamos da esperança matemática 3.2.1 observamos os valor esperados quando realizados determinadas quantidade de experimentos n, ou seja, obtemos em resumo uma média desses dados. A função da variância num estudo assim, é mostrar quão dispersos estão os valores singulares em relação a este valor esperado, dentro disso o estudo da covariância mostra como se comporta essa relação, de forma simples se resultar em valor negativo os valores de X e Y se condicionam opostos, desta forma, os valores, os valores menores correspondem aos maiores e os maiores a valores menores e se o resultado for positivo teremos que, valores menores se ligam a valores menores e de mesmo modo os maiores.

3.3 PIB Araguaína-TO

Esse indicador econômico foi criado para entender o impacto populacional sobre a produtividade. Inicialmente, esse conceito foi dado por Simon Kuznets em 1930 nessa finalidade de entender a economia mundial, mas, efetivamente apenas em 1953 foi feito o primeiro cálculo do PIB de um território nacional pelas Nações Unidas.

Segundo[14]:

Definição 3.5. PIB. *Soma de todos os bens e serviços finais produzidos por um país, estado ou cidade, geralmente em um ano.*

Entende-se segundo o estudo do PIB, que quanto maior é esse indicador maior são as atividades econômicas da região. Assim ele representa o quanto a economia produziu num certo período de tempo, essa função é atribuída ao IBGE que organiza, calcula e compartilha essas demandas.

A principal composição do PIB Araguainense é dado pelos setores agropecuário, industrial e serviços. Os dados abaixo foram organizados pela Secretaria do Planejamento e Orçamento (SEPLAN) segundo censo do IBGE.

Tabela 3.1: Valores em real por setor de atividade no PIB Araguaína-TO

Ano	Agropecuário (1000)	Indústria (1000)	Prestação de serviços (1000)	PIB (1000)
2008	23.944,99	215.509,05	1.031.041,60	1.466.466,70
2009	30.466,67	274.152,29	1.192.294,54	1.695.273,21
2010	35.033,54	309.754,84	1.397.210,08	1.980.667,65
2011	35.719,61	309.385,26	1.535.367,03	2.139.827,56
2012	37.497,78	333.638,52	1.782.498,88	2.434.281,10
2013	37.313,83	405.125,48	2.111.909,22	2.887.138,78
2014	46.971,01	402.454,77	2.247.378,95	3.053.584,53

Fonte - Elaborado a partir dos dados organizados por [17]

Observamos então na tabela, que as principais contribuições para o avanço do PIB Araguainese no intervalo dos anos de 2008-2014 se encontram destacados. Dessa forma, nosso estudo terá base nesses dados (agropecuário, prestação de serviços e indústrias) no período já colocado.

3.4 Correlação

A análise do coeficiente de correlação tem como objetivo medir quão forte, ou fraca, é a relação linear entre duas variáveis e o que elas representam, a partir da avaliação do grau de associação entre duas variáveis X e Y. Através dele sabemos se alterações que ocorrem em uma variável ocasiona alteração na outra, mostrando se existe uma ligação de dependência entre elas.

Como ressalta[16], o cargo do coeficiente de correlação é medir em específico um tipo de dependência recíproca, a interdependência linear. Sendo assim, se a relação entre duas variáveis X e Y se caracterizarem ao tipo não-linear o valor em módulo do coeficiente de correlação não será grande.

O coeficiente de correlação linear entre duas variáveis X e Y é representado por R_{xy} e dado pela fórmula:

$$R_{xy} = \frac{\displaystyle\sum_{i=1}^{n} x_i y_i - n.\overline{x.y}}{\sqrt{(\displaystyle\sum_{i=1}^{n} x_i^2 - n\overline{x}^2)}\sqrt{(\displaystyle\sum_{i=1}^{n} y_i^2 - n\overline{y}^2)}} \tag{3.4}$$

sendo $x_1, x_2, ..., x_n$ e $y_1, y_2, ..., y_n$ os valores medidos das variáveis e, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ e $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Desta forma temos:

$$
\begin{aligned}
R_{xy} &= \frac{\sum x_i.y_i - n\frac{\sum x_i}{n}\frac{\sum y_i}{n}}{\sqrt{(\sum x_i^2 - n\overline{x}^2)}\sqrt{(\sum y_i^2 - n\overline{y}^2)}} \\
&= \frac{n}{n}\frac{(\sum x_i.y_i - \frac{\sum x_i.\sum y_i}{n})}{\sqrt{\sum x_i^2 - n\overline{x}^2}\sqrt{\sum y_i^2 - n\overline{y}^2}} \\
&= \frac{n\sum x_i.y_i - \sum x_i \sum y_i}{\sqrt{n}^2[(\sum x_i^2 - n\overline{x}^2)(\sum y_i^2 - n\overline{y}^2)]} \\
&= \frac{n\sum x_i.y_i - \sum x_i.\sum y_i}{\sqrt{n}[\sum x_i^2 - n(\frac{\sum x_i}{n})^2]n[\sum y_i^2 - n(\frac{\sum y_i}{n})^2]} \\
&= \frac{n\sum x_i.y_i - \sum x_i.\sum y_i}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}}
\end{aligned}
$$

Assim, temos:

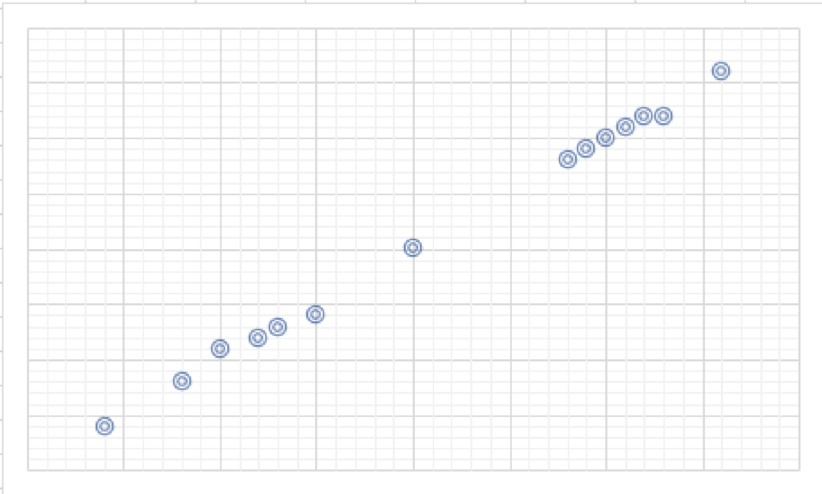
$$
R_{xy} = \frac{n\sum x_i.y_i - \sum x_i.\sum y_i}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}} \tag{3.5}
$$

Por meio desta fórmula é possível identificar se existe ou não uma relação entre duas variáveis quaisquer. Onde $-1 \leq R_{xy} \leq 1$.

- $R_{xy} = 1$, a correlação é perfeita positiva.
- $R_{xy} = 0$, a correlação é nula.
- $R_{xy} = -1$, a correlação é perfeita negativa.
- +1 e -1 indicam se a correlação é positiva ou negativa, e o sinal a direção linear.

Tem-se que a relação entre duas variáveis X e Y pode representada num gráfico de dispersão, dessa forma, se estes valores caracterizarem uma reta afirma-se ter uma relação linear. Pode-se definir essa relação em três formatos, observe as figuras seguintes:

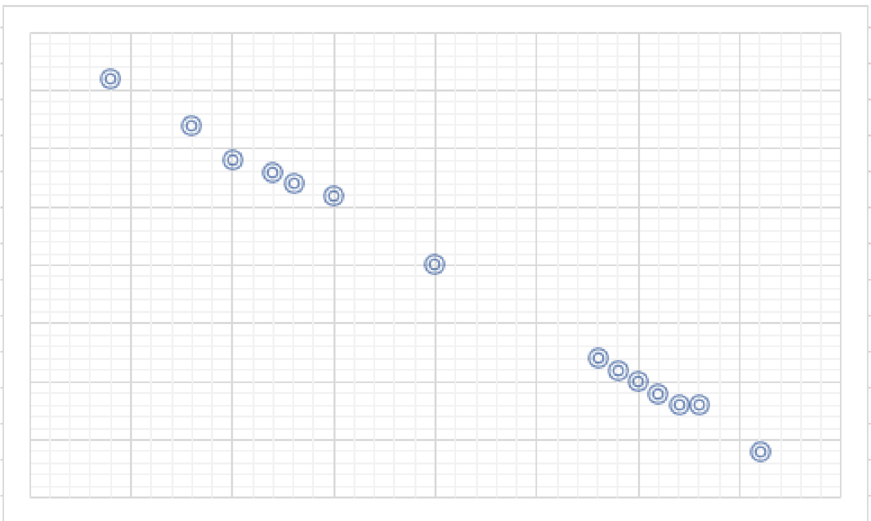
Gráfico 01 - $0 < R_{xy} < 1$



Fonte: Construção própria.

Nesse gráfico é possível observar que os valores em Y mantém uma relação de dependência com os de X, de tal modo que o valor da sua correlação muito se aproxima de 1, caracterizando uma reta crescente.

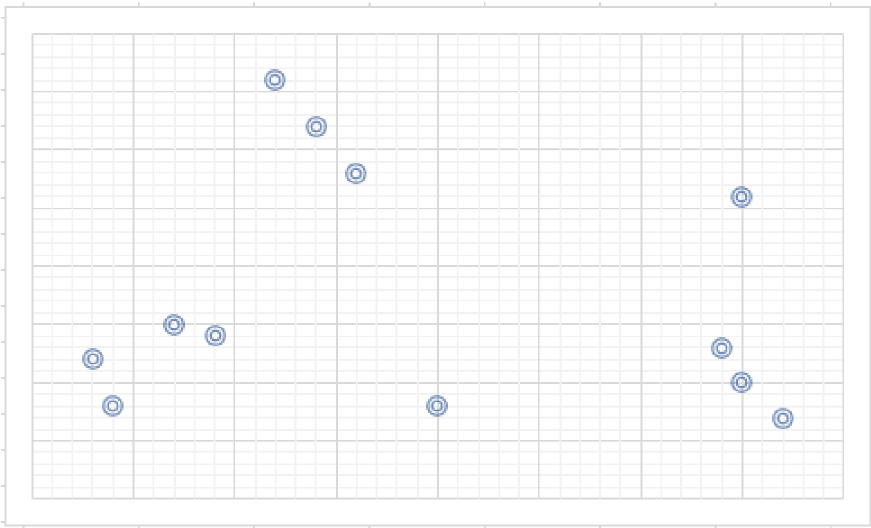
Gráfico 02 - $-1 < R_{xy} < 0$



Fonte: Construção própria.

Aqui pode-se perceber o mesmo do gráfico anterior, a correlação está próxima de -1, o que diz, que esse valores estão mantendo uma relação de dependência linear.

Gráfico 03 - $R_{xy} \approx 0$



Fonte: Construção própria.

Aqui se verifica valores dispersos, declarando que não existe uma dependência entres as variáveis X e Y. Sendo assim, o valor da correlação é próximo de zero.

Como mostrado nos gráficos de dispersão, os valores (x_i, y_i) plotados em um gráfico poderá descrever visualmente o quão associados eles estão.

Associando as variáveis agropecuário X e PIB anual Y da Tabela 5.8

Tabela 3.2: Valores em real por setor de atividade no PIB Araguaína-TO

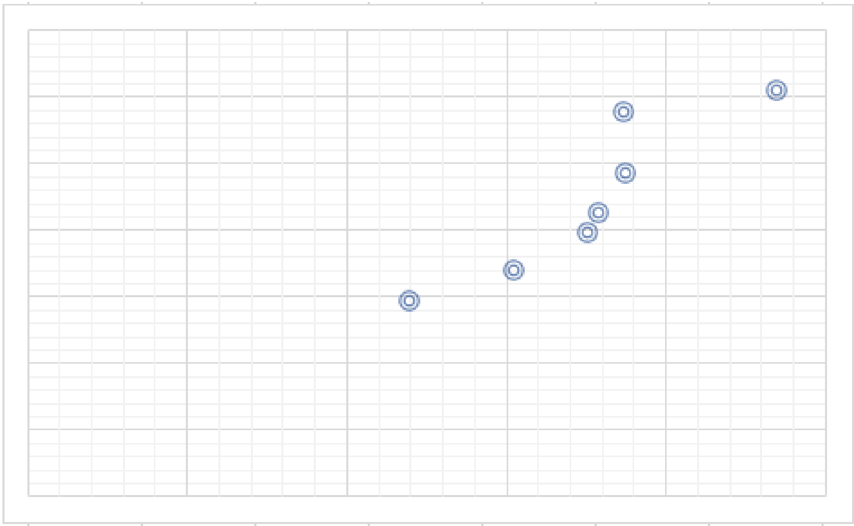
Ano	Agropecuário (1000)	PIB (1000)
2008	23.944,99	1.466.466,70
2009	30.466,67	1.695.273,21
2010	35.033,54	1.980.667,65
2011	35.719,61	2.139.827,56
2012	37.497,78	2.434.281,10
2013	37.313,83	2.887.138,78
2014	46.971,01	3.053.584,53

Fonte - Elaborado a partir dos dados organizados por [17]

Verifica-se em 3.2 que, dentro período entre os anos de 2008-2014 o setor agropecuário possibilitou valores que contribuíram para concretização do PIB Araguainense, por isso sua colocação na tabela juntamente dos valores finais dessa produção do PIB.

Pegando os valores de 3.2 podemos produzir um diagrama de dispersão para uma leitura gráfica. Dessa forma:

Gráfico 04 - Diagrama de dispersão: produção setor de atividade (X) e PIB (Y)



Fonte: Construção própria.

Observe que o diagrama mostra uma relação de dependência entre as variáveis, pois seus valores se aproximam da imagem de uma reta. Para verificar essa veracidade iremos se valer da fórmula, para efetuar os cálculos:

$$R_{XY} = \frac{n\sum x_i.y_i - \sum x_i.\sum y_i}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}}$$

$$R_{xy} = \frac{1,58679 * 10^{11}}{\sqrt{72.083.475.112,03} * \sqrt{714.635.238.490.085,00}}$$

Colocando o valores e aproximação para melhoria dos cálculos temos:

$$R_{xy} = \frac{1,59 * 10^{11}}{1,74 * 10^{11}}$$

$$R_{xy} = 0,908711552$$

Desta forma compreende-se que os valores mantém uma associação linear dado que o valor da correlação que define isso de aproxima de 1, ou seja, mantém uma relação positiva quase perfeita. O que pode ser observado tanto pela forma como os valores de (x_i, y_i) se associam no diagrama de dispersão, e pelo cálculo de correlação a existência de uma relação de dependência entre estes valores, o que diz que o PIB depende da produção financeira do Agropecuário.

Capítulo 4

MODELOS DE REGRESSÃO

4.1 Regressão Linear Simples

A análise de regressão tem o intuito de selecionar um conjunto de dados amostrais e através dele achar uma equação que defina a relação existente entre as variáveis.

Esse modelo mostra por meio de uma equação a relação linear existente entre a variável reposta Y e a variável regressora X, e quando plotadas num gráfico chamamos de reta de regressão. Esta reta de regressão é a que melhor representa a relação existente entre as variáveis com o menor erro possível, ou seja, se olharmos para uma equação no gráfico em formato de reta veremos a que melhor se aproxima de todos os pontos plotados no gráfico.

Para definirmos a equação de regressão, iremos considerar duas variáveis aleatórias quantitativas X e Y. Sabemos que a $E(Y|x)$ é a esperança de Y condicionada a x dado que $X = x$. Analogamente, uma definição correspondente vale para $E(X|y)$, que é função de y. Considerando aqui o caso em que X e Y seja a população, modelamos:

$$Y_i = \mu_i + \epsilon_{ij}, \tag{4.1}$$

$i = 1, 2, ..., n$ e $j = 1, 2..., m$, onde μ é a média das n observações.

Para $\mu(x)$ uma função de X, podemos considerar que X e Y mantêm uma distribuição normal conjunta bidimensional, nesse fato, $\mu(x)$ e $\mu(y)$ são funções lineares. Para tal, aderimos ao modelo $E(Y|x)$ que é:

$$E(Y|x) = \mu(x) = \alpha + \beta.X \tag{4.2}$$

alternativamente temos,

$$Y_{ij} = \mu(x_i) + \epsilon_{ij} \Rightarrow Y_i = \mu(x_i) + \epsilon_i \tag{4.3}$$

Reescrevendo esse modelo temos:

$$Y_i = E(Y|x_i) + \epsilon_i = \alpha + \beta.x_i + \epsilon_i, i = 1, 2, ..., n \tag{4.4}$$

Sendo assim a equação para regressão linear é:

$$Y_i = \alpha + \beta.X_i + \epsilon_i$$

(população),

$$Y_i = a + b.X_i + \epsilon_i$$

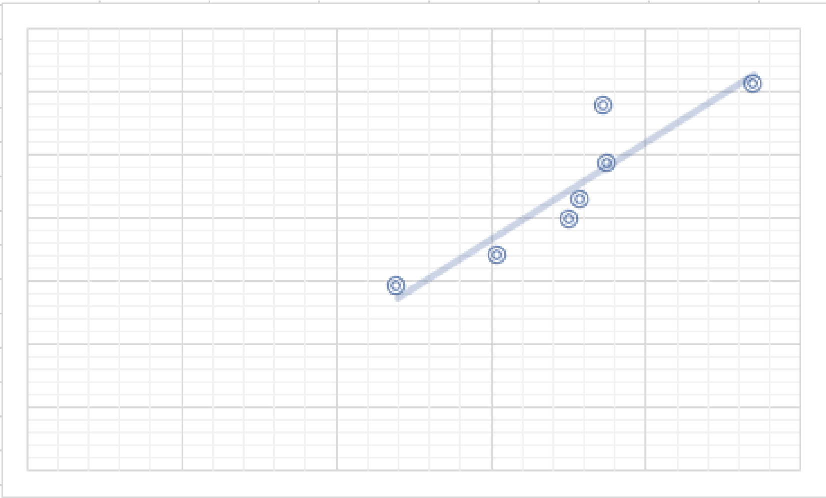
(amostra).

Admitindo o estudo pelo modelo linear, segundo [15]:

- A relação linear entre X e Y é linear;
- Os valores X são fixos, isto é, X não é uma variável aleatória;
- A média do erro é nula, isto é, $E(\epsilon) = 0$;
- Para um dado valor de X, a variância do erro ϵ sempre é θ^2, denominada variância residual, isto é,
$$E(\epsilon^2) = \theta^2 \Rightarrow E[Y_i - E(Y_i|Y_i)]^2 = \theta^2; \tag{4.5}$$
- O erro de uma observação é não-correlacionado com o erro em outra observação, isto é, $E(\epsilon_i \epsilon_i) = 0$, com $i \neq j$;
- Os erros têm distribuição normal.

O erro no modelo de regressão mostra como os valores se distanciam da linha de tendência, ou seja, a distância média desses valores comparando-os a linha de regressão formada. Quão menor for esses valores, menor será a distância desses valores da linha formada.

Gráfico 05 - Produção agrícola em 2008-2014 (X) e PIB 2008-2014 (Y)

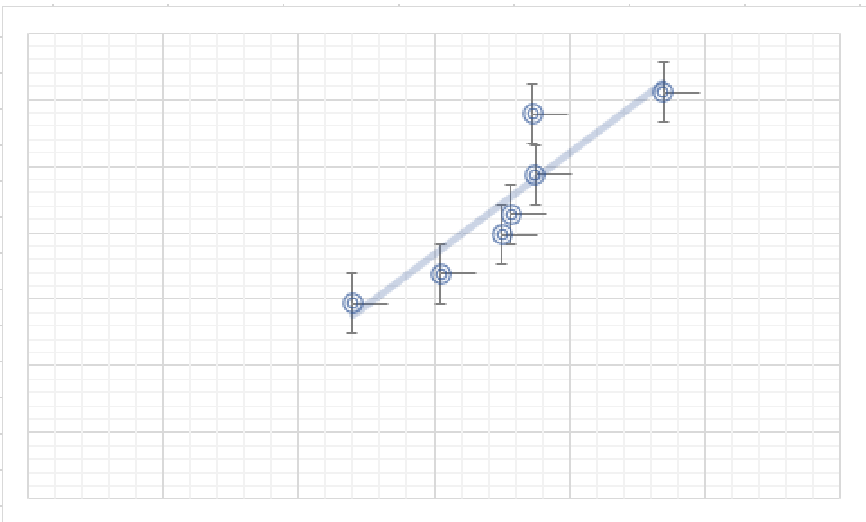


Fonte: Construção própria.

A linha de tendência traçada é uma 'média' por assim dizer, do resumo desses valores dispersos. Se imaginarmos uma reta mensurando a distância de cada ponto à linha traça obteremos o erro com relação a linha de tendência. Dessa forma podemos dizer pelos dados gráficos que o erro é exatamente $\epsilon = Y_i - (aX + b)$.

Adicionando o erro padrão em cima do diagrama de dispersão formado, observe o (Gráfico 06).

Gráfico 06 - Produção agrícola em 2008-2014 (X) e PIB 2008-2014 (Y)



Fonte: Construção própria.

As linhas representa o erro ϵ_i para mais e para menos, os valores que estão exatamente em cima da reta esperada não têm erro, os valores que se mantém distantes da reta, obterão um valor para esse erro. Em resumo, é possível compreender pela leitura dos gráficos acima, que existe um erro (ϵ_i), em relação a linha de tendência formada. Esse erro mostra a existência de uma grande variabilidade na amostra pois para cada valor uma equação diferente, assim temos os valores estimados, a é um estimador de α e b um estimador de β e caso fosse utilizado todo o conjunto de população $a = \alpha$ e $b = \beta$.

4.2 Estimando Parâmetros

Para introduzirmos esta seção, consideremos algumas conjecturas sobre as variáveis aleatórias envolvidas. Uma suposição sobre X (variável regressora) é que esta não está sujeita a variação, desta forma podemos entender que X é fixa ou determinística.

Outra suposição sobre esta variável é que, para cada valor de x de X os erros estarão dispersos envolto da média $\alpha + \beta.x$ com média zero:

$$E(\epsilon_i|x) = 0 \tag{4.6}$$

Uma terceira suposição, é considerar que esses erros tenham variabilidades iguais ao redor dos níveis de X, de forma que:

$$Var(\epsilon_i|x) = \theta^2 \tag{4.7}$$

E por fim, iremos supor restrição na qual os erros (ϵ_i) não mantenham algum tipo de relação. Então seja uma amostra de n observações, obtemos n pares (X_i, Y_i), $i = 1, 2, ..., n$ de modo a satisfazer o módulo:

$$Y_i = \alpha + \beta.X_i + \epsilon_i, i = 1, 2, ..., n \tag{4.8}$$

Para encontrar α e β precisamos de um critério. Dessa forma, iremos adotar valores de α e β que minimizem a soma dos quadrados dos erros. Esse método considera uma reta com n pontos $(x_1, y_1), ..., (x_n, y_n)$, analisando a diferença entre cada valor de Y_i com a reta, que concorda com o valor de X_i. Dando origem a uma reta que demonstra a menor soma de quadrados destas diferenças.

$$\epsilon_i = Y_i - (\alpha + \beta_i), i = 1, 2, ..., n \tag{4.9}$$

Assim, conseguimos obter a quantidade de informações perdidas na soma dos quadrados.

$$SQ(\alpha, \beta) = \sum \epsilon_i^2 = \sum [Y_i(\alpha + \beta.X_i)]^2 \tag{4.10}$$

Dessa forma, para cada valor de α e β será obtido um resultado diferente nesta soma dos quadrados (SQ). Para tal, deve-se encontrara o mínimo de uma função para as variáveis α e β, o que é necessário a derivada (consulte[18]) em relação a α e β igualando a zero.

Observe que para dado valor α e β, obtemos um resultados na soma de quadrados. E para encontrar a forma mínima da soma, usaremos solução dos mínimo quadrado(MQ). Pelas derivadas parciais (consulte[18]) as soluções de α e β devem satisfazer:

$$\begin{cases} n\alpha + \beta \sum X_i = \sum Y_i \\ \alpha \sum X_i + \beta \sum .X_i^2 = \sum X_i Y_i \end{cases} . \tag{4.11}$$

O que implica na solução:

$$\beta = \frac{\sum_{i=1}^{n} X_i.Y_i - nX.Y}{\sum_{i-1}^{n} X_i^2 - nX_i^2} \tag{4.12}$$

e,

$$\alpha = Y_i - \beta X_i \tag{4.13}$$

Por essas equações temos os valores ajustados de α e β de forma que minimizem a soma dos quadrados da diferença entre a reta e os pontos da amostra, diminuindo assim os resíduos.

4.3 Regressão Linear Múltipla

Muitos problemas admitem que previsões podem ser melhor explicadas quando se está em função de duas ou mais variáveis. O modelo de regressão linear múltipla tem que a variável Y, é função de X_k $(k = 1, 2, ..., n)$ variáveis independentes ou explicativas, dessa forma esse modelo estabelece uma equação para predizer valores de Y levando em conta as diversas variáveis X_k.

Como nesse estudo consideramos n observações para cada variável independente, temos que para cada valor $X_1, X_2, ..., X_k$ fixos Y_i é variável aleatória. Tal modelo estatístico para k variáveis independentes é escrito como:

$$Y_i = b + ax_i 1 + aX_i 2 + ... + aX_{ik} + \epsilon_i \tag{4.14}$$

de forma temos que:

- Y_i= valores observados de Y;
- b, a_1,...,a_k= parâmetros a serem estimados;
- X_i1,...,$X_i k$ i-ésimo nível das k variáveis independentes $(i = 1, 2, ..., n)$;

- ϵ_i= erro.

Consideremos esse modelo de regressão linear com duas variáveis regressoras:

$$Y_i = b + a_1 b_i 1 + a_2 b_i 2 + \epsilon_i \tag{4.15}$$

Visando minimizar o erro, usaremos o método dos mínimos quadrados (MMQ), para obtenção da equação estimada:

$$\sum .\epsilon_i^2 = \sum (Y_i - \beta - \beta_1 .X_{1i} - \beta_2 .X_{2i}) \tag{4.16}$$

Assim, os valores estimados de β, β_1 e β_2 que minimizam os erros é dados pela solução do sistema, após aplicação de derivada para cada um dos coeficientes:

$$\begin{cases} n.\hat{\beta} + \hat{\beta}_1.(\sum .x_1) + \hat{\beta}_2.(\sum .x_2) = \sum .y \\ \hat{\beta}.(\sum .x_1) + \hat{\beta}_1.(\sum .x_1^2) + \hat{\beta}_2.(\sum .x_1.x_2) = \sum x_1 y \\ \hat{\beta}.(\sum .x_2) + \hat{\beta}_1.(\sum .x_1.x_2) + \hat{\beta}_2.(\sum .x_2^2) = \sum .x_2.y \end{cases} \tag{4.17}$$

Esse sistema resulta nos valores estimados que minimizam a soma dos resíduos (SSE), reescrevendo abaixo:

$$\hat{\beta}_1 = \frac{(\sum .x_1^2 - \frac{(\sum .x_1)^2}{n})(\sum .x_2.y - \frac{\sum .x_2.\sum .y}{n}) - (\sum .x_1.x_2 - \frac{\sum .x_1.\sum .x_2}{n}).(\sum .x_1.y - \frac{\sum .x_1.\sum .y}{n})}{(\sum .x_1^2 - \frac{(\sum .x_1)^2}{2}).(\sum .x_2 - \frac{(\sum .x_2)^2}{n}) - (\sum .x_1.x_2 - \frac{\sum .x_1.\sum .x_2}{2})^2} \tag{4.18}$$

$$\hat{\beta}_2 = \frac{(\sum .x_1.y - \frac{\sum .x_1.\sum .y}{n}) - \hat{\beta}_1(\sum .x_1.x_2 - \frac{\sum .x_1.\sum .x_2}{n})}{\sum .x_1^2 - \frac{(\sum .x_1)^2}{n}} \tag{4.19}$$

$$\hat{\beta} = \overline{y} - \hat{\beta}_2.\overline{x_1} - \hat{\beta}_1.\overline{x_2} \tag{4.20}$$

Logo, temos a equação do modelo de regressão ajustado dado por:

$$\hat{Y} = \hat{\beta} + \hat{\beta}_1.X_1 + \hat{\beta}_2.X_2 \tag{4.21}$$

4.4 Qualidade do Ajustamento

A equação da reta de regressão é uma das formas de se explicar a variação de Y resultante das variações de X_i, com $i = 1, 2, ..., n$. Para mensurarmos a qualidade que se tem como o modelo de regressão dado, nos valemos de R^2. O coeficiente de determinação, mais conhecido por R^2 tem a função de explicar a qualidade do ajustamento, ou seja, o quão explicativo é o modelo formado. Esse coeficiente é uma medida descritiva que calcula a quantidade de variabilidade explicada no ajustamento de reta, dado pelo modelo de regressão.

Essa ferramenta estatística considera:

- $SST = \sum .(y_i - \overline{y})$ Soma dos quadrados totais

Essa soma, é a dispersão dos valores em torno da média dos valores da amostra $\overline{y}$ - variação total.

- $SSE = \sum .(y_i - \hat{y})$ Soma dos quadrados dos resíduos

Já essa soma, leva em conta a dispersão em tono da equação de regressão $\hat{y}$ - variação não explicada.

Assim, temos:

$$SST = SSE + SSR \tag{4.22}$$

O mesmo que:

$$\sum(y_i - \overline{y})^2 = \sum(y_i - \hat{y}_i)^2 + \sum(\hat{y}_i - \overline{y})^2 \tag{4.23}$$

Dessa forma, o quociente entre SSR e SST implica na medida da variação total que é explicada pelo ajustamento feito.

$$r^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} \tag{4.24}$$

Considerando a fórmula acima, o resultado dá origem a R^2. Esse coeficiente tem sua variação $0 \leq r^2 \leq 1$, se o valor do coeficiente se aproxima de 1, dizemos que o ajustamento explica bem os valores da amostra, se mais próximo de 0 dizemos que esse ajustamento tem um poder explicativo menor.

Em outras palavras, a qualidade do ajustamento pode ser definida para dois momentos:

$r^2 \approx 0 \rightarrow$ modelo linear muito pouco adequado;

$r^2 \approx 1 \rightarrow$ modelo linear bastante adequado.

São essas as considerações que faremos para analisar se o modelo de regressão encontrado é adequado para explicar os valores da amostra ou não.

Capítulo 5

APLICAÇÕES

Os dados que serão usados aqui foram arranjados pela SEPLAN, segundo censo do IBGE do PIB municipal no período de 2008-2014, concentraremos esforços nas informações dos setores que mais contribuem na produção final do PIB do município de Araguaína-TO como mostra a tabela 5.8.

A amostra é constituída pelos três setores que mais fomentam no cálculo do PIB, que são eles: setor de indústrias, setor agropecuário e setor de prestação de serviços. Serão apresentados dois conjuntos de estudos, no primeiro estudo foi utilizado o modelo de regressão linear simples, e no segundo considerámos o modelo de regressão linear múltipla.

Caracterizamos como método científico de estudo, o método dedutivo, onde "toda a informação ou conteúdo fatual da conclusão já estava, pelo menos implicitamente, nas premissas"[8]. A finalidade aqui é "a descrição das características de determinada população ou fenômeno ou, então, o estabelecimento de relações entre variáveis"[19], compreendendo assim o conjunto da população denominada PIB e a relação que mantém com as variáveis das amostras.

5.1 Estudo 1 - Análise de regressão linear simples

Aspirando compreender se os setores citados influenciam na produção final do cálculo do PIB no município, iremos ordenar as seguintes situações:

- PIB Y em função do Agropecuário X;
- PIB Y em função da Indústria X;
- PIB Y em função da Prestação de serviços X.

Discorreremos esses subestudos separadamente, para contemplar esta secção sobre o modelo de regressão linear simples.

5.1.1 Análise de regressão linear simples - Agropecuário e PIB

Colocando novamente 3.2:

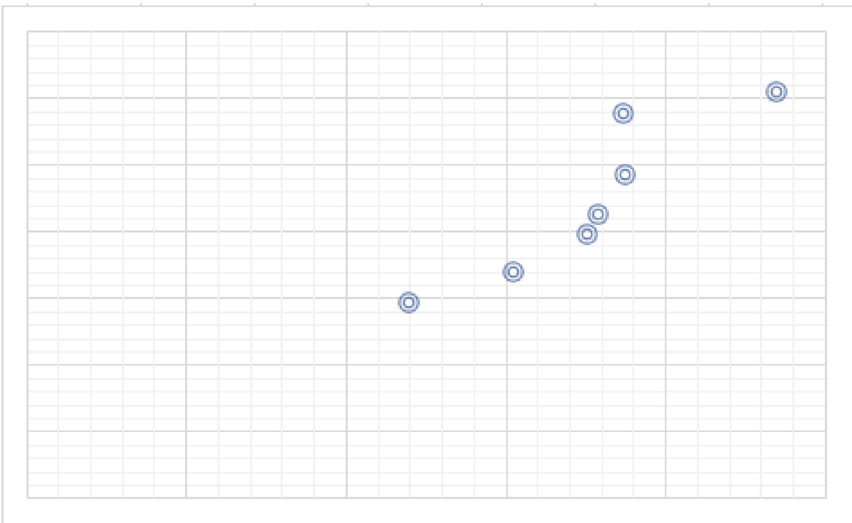
Tabela 5.1: Valores em real por setor de atividade no PIB Araguaína-TO

Ano	Agropecuário (1000)	PIB (1000)
2008	23.944,99	1.466.466,70
2009	30.466,67	1.695.273,21
2010	35.033,54	1.980.667,65
2011	35.719,61	2.139.827,56
2012	37.497,78	2.434.281,10
2013	37.313,83	2.887.138,78
2014	46.971,01	3.053.584,53

Fonte - Elaborado a partir dos dados organizados por [17]

A tabela 5.1 mostra os valores da colaboração do setor agropecuário no PIB Araguainense. Inicialmente iremos observar essa relação no gráfico em que o valor para o eixo X é o agropecuário e no eixo Y o PIB, no período já colocado para identificar se a relação entres eles é de fato linear.

Gráfico 07 - Diagrama de dispersão: Agropecuário (X) e PIB (Y)



Fonte: Construção própria.

Podemos observar no gráfico acima que os valores mostram um certo grau de relação entre si, dessa forma nos insinuam o cálculo da correlação para identificarmos o grau de relação que eles mantém. Como já visto na secção 3.4 por meio de 3.5 temos que:

$$R_{xy} = \frac{1,59^{11}}{1,74 * 10^{11}}$$

$$R_{xy} = 0,908711552$$

O que identifica uma relação linear entre as variávei X e Y correspondentes, ou seja, setor agropecuário e PIB se relacionam de modo que a produção dos valores no agropecuário (variável independente) provocam variações no PIB (variável dependente). Se isso ocorre podemos encontrar a equação da reta de regressão que melhor explica esses valores.

Por 4.12 temos que achar o valor de β ajustados, assim:

$$\beta = \frac{575.028 - 7.(789.084.706,70)}{900.950.118,5 - 7.(124.455.169,80)}$$
$$\beta = \frac{22.668.462.742,94}{29.763.929,9}$$
$$\beta = 76,16$$

Agora por 4.13:

$$\alpha = \overline{Y} - 76,16.\overline{X}$$

$$\alpha = 2.236.748,50 - 2.686.787,712$$

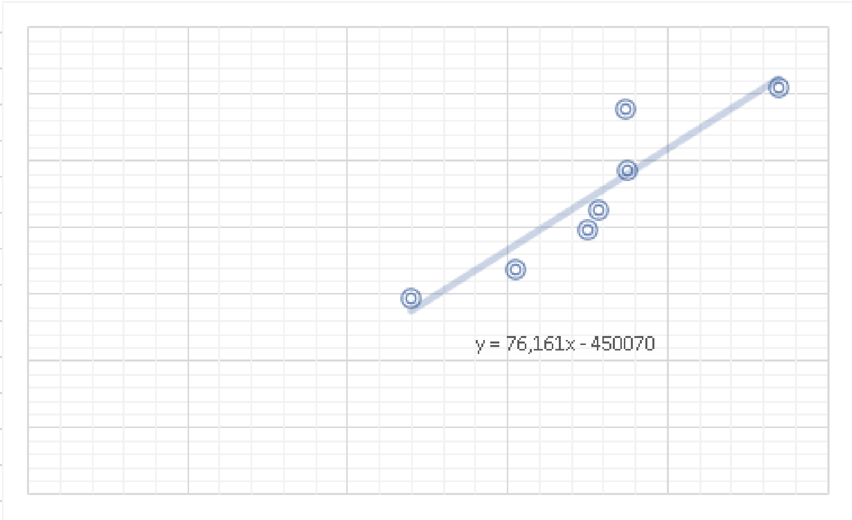
$$\alpha = 450070$$

Notamos então que nossa equação com parâmetros ajustados é:

$$Y = 76,16X - 450070$$

Visualizando no gráfico, conseguimos perceber melhor essa relação. E por 4.4 conseguimos o valor de $R^2 = 0,75$, que nos diz que estes valores estão bem explicados pela equação da reta.

Gráfico 08 - Diagrama de dispersão: Agropecuário (X) e PIB (Y)



Fonte: Construção própria.

Agora iremos substituir os valores de X para as seguintes situações: $X_1 = 15.000,00$, $X_2 = 22.900,00$ e $X_3 = 26.900,99$. Faremos uma breve leitura do seu significado.

Para X_1, temos:

$$Y_1 = 76,16.(15.000,00) - 450070$$

$$Y_1 = 692.330$$

Esse valor corresponde ao PIB quando as contribuição do agropecuário estiver em $15.000,00$, percebemos que esse valor é menor que o valor calculado no ano de 2008, ou seja, com a queda da contribuição desse setor o PIB também sofrerá.

Em $X_2 = 22.900,00$, temos:

$$Y_2 = 76,16.(22.900,00) - 450070$$

$$Y_2 = 1.293.994$$

Esse valor ainda é menor que o PIB marcado no ano de 2010 pois a produção do setor agropecuário ainda está abaixo do ano referência, ou seja ainda mostra baixa no PIB.

Para $X_3 = 26.900,99$:

$$Y_3 = 76,16.(26.900,99) - 450070$$

$$Y_3 = 1.598.639$$

Perceba que esse valor de X se encontra acima do valor no ano de 2010, o que nos diz por dedução que o valor para o PIB quando $X = 26.900,99$ estará também entre nesse intervalo de tempo, como vimos, de fato está no intervalo. Conseguimos entender aqui o setor Agropecuário tem grande interferência no cálculo do PIB, pois, suas alterações causam mudanças notáveis nesta soma.

Tabela 5.2: Valores em real por setor de atividade no PIB Araguaína-TO

X_i	Agropecuário (1000)	PIB (1000)
X_1	15.000,00	692.330
X_2	22.900,00	1.293.994
X_3	26.900,99	1.598.639

Fonte - Produção própria

Dessa forma, é nítido perceber que quanto maior o valor da produção do setor agropecuário maior será a soma total do PIB nessa situação, e quanto menor for esse valor menor o valor final do PIB no município, o que mostra uma influência significativa desse setor levando em conta que o PIB é soma de todos os bens financeiros produzidos pelo município.

5.1.2 Análise de regressão linear simples - Indústria e PIB

Construindo a tabela referente aos valores do estudo.

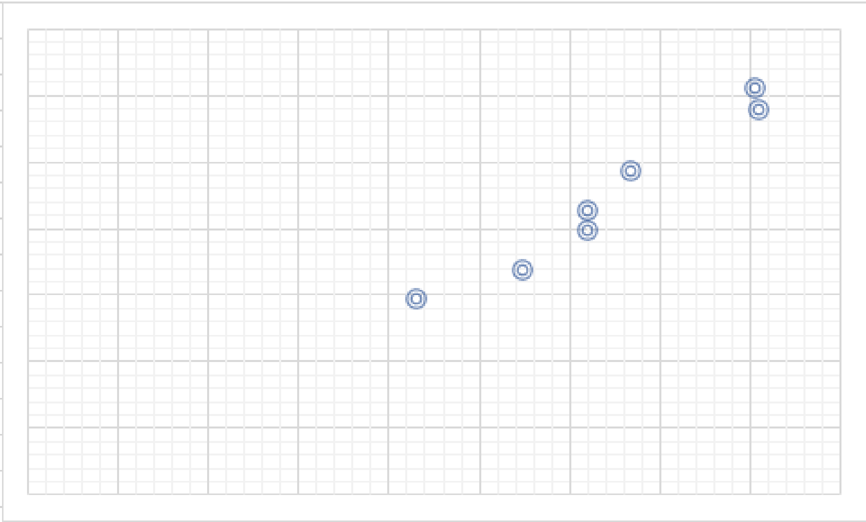
Tabela 5.3: Valores em real por setor de atividade no PIB Araguaína-TO

Ano	Indústria (1000)	PIB (1000)
2008	215.509,05	1.466.466,70
2009	274.152,29	1.695.273,21
2010	309.754,84	1.980.667,65
2011	309.385,26	2.139.827,56
2012	333.638,52	2.434.281,10
2013	405.125,48	2.887.138,78
2014	402.454,77	3.053.584,53

Fonte - Elaborado a partir dos dados organizados por [17]

Inicialmente faremos dos valores da tabela 5.3 um diagrama de dispersão, para analisarmos se estes mantêm uma relação de fator linear.

Gráfico 09 - Diagrama de dispersão: Indústria (X) e PIB (Y)



Fonte: Construção própria.

De fato, é notável a forma como estes valores se relacionam de forma alinhada tendendo a formar uma reta, o que caracteriza uma relação de interdependência linear entres eles. Dessa forma, podemos nos certificar por meio de cálculos matemáticos se isso realmente acontece efetuando os cálculos do coeficiente de correlação.

$$R_{xy} = \frac{(5,2611.10^{12}) - (2250020,15.15657239,53)}{\sqrt{7(5,81327.10^{12}.5,06259.10^{12})}\sqrt{7(2,8226.10^{14}.2,4519.10^{14})}}$$

$$R_{xy} = 0,98$$

Esse coeficiente determina uma relação linear positiva quase perfeita entre os valores da produção do setor de Indústrias e o PIB, expondo que as alterações feitas nos valores de X_i afetarão a soma final do PIB da cidade. Como isso ocorre, podemos calcular a equação que determina a reta de regressão para esse caso.

Por 4.12 é possível achar o valor de β ajustado:

$$\beta = \frac{5,26711.10^{12} - 7(309754,84.2139827,56)}{5,81327 - 7(95948060903)}$$

$$\beta = 8,53$$

Agora por 4.13 calculamos o valor ajustado de α

$$\alpha = 2.139.827,56 - (8,563.309754,84)$$

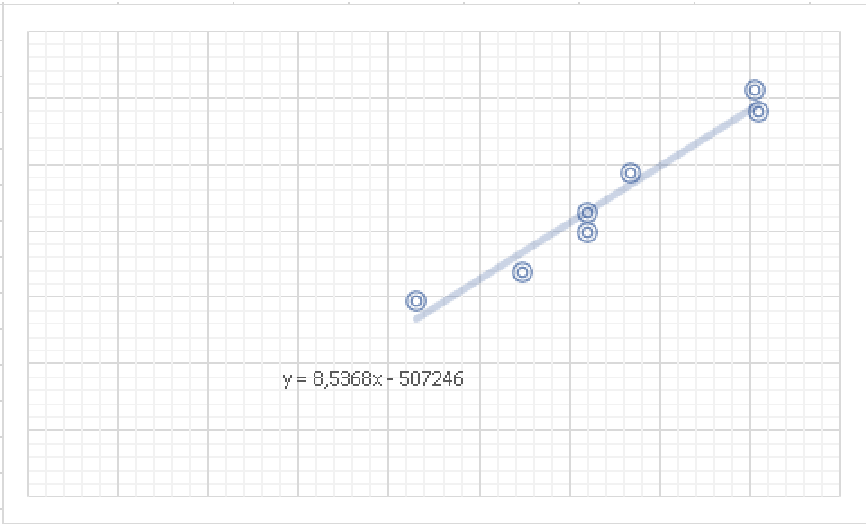
$$\alpha = 507246$$

Logo a equação que melhor representa os valor da tabela 5.3 é:

$$Y = 8,53X - 507246$$

Graficamente, observamos como a reta de regressão e essa equação que define os valores da reta. Como temos $R^2 = 0,96$ por 4.4, conseguimos entender que essa equação possui um alto poder explicativos para estes valores da amostra tabelados acima.

Gráfico 10 - Diagrama de dispersão: Indústria (X) e PIB (Y)



Fonte: Construção própria.

Agora alguns estudos podem ser montados para identificarmos uma melhor leitura dos resultados. Assim iremos considerar alguns valores para X: $X_1 = 200.000,00$, $X_2 = 300.000,00$ e $X_3 = 400.000,00$.

Para X_1, temos:

$$Y_1 = 8,53(200.000,00) - 507246$$

$$Y_1 = 1.202.754$$

O valor de X_1 está abaixo do valor no ano 2008, quando colocado na equação da reta o valor de Y_1 ficou abaixo do valor PIB neste ano, o que diz da relação de independência entre eles.

Para X_2, temos:

$$Y_2 = 8,53(300.000,00) - 507246$$

$$Y_2 = 2.057.754$$

Note que X_2 está abaixo do valor no ano de 2010, e Y_2 correspondente está acima do valor no PIB neste ano, o que diz que esse valor sofre interferência de outras variáveis do PIB, sabemos que o PIB é soma destes e outros setores de atividades econômicas da cidade. Entendemos então que, se houver uma queda mínima no setor de indústrias não irá interferir tanto considerando os valores dos outros setores.

Para X_3, temos:

$$Y_3 = 8,53(400.000,00) - 507246$$

$$Y_3 = 2.912.754$$

Aqui ocorre novamente a situação de X_2, a queda nesse setor não irá influenciar em uma queda no PIB, levando em conta a produção dos demais setores.

Tabela 5.4: Valores em real por setor de atividade no PIB Araguaína-TO

X_i	Indústria(1000)	PIB (1000)
X_1	200.000,00	1.202.754
X_2	300.000,00	2.057.754
X_3	400.000,00	2.912.754

Fonte - Produção própria

Note que, as baixas no setor não terão tanta influência para uma queda no PIB do município, mas, considerando a alta no PIB mesmo com a diminuição desse setor inferimos que quando estiver em alta terá aumento significativo no PIB. Aqui podemos deduzir ou supor por exemplo que na cidade não há muitas atividades de Indústria, pelo fato de uma movimentação tanto exporadica. São suposições que podem ser feitas para estudo mais aprofundados, tida com os valores encontrados com essa pesquisa.

5.1.3 Análise de regressão linear simples - Prestação de serviços e PIB

Como nas situações anteriores, a amostra em questão para estudo é a seguinte tabela.

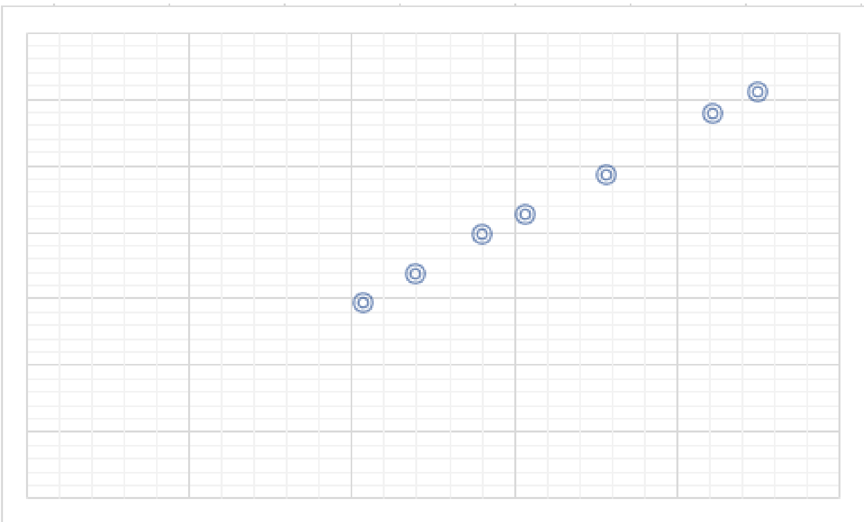
Tabela 5.5: Valores em real por setor de atividade no PIB Araguaína-TO

Ano	Prestação de serviços (1000)	PIB (1000)
2008	1.031.041,60	1.466.466,70
2009	1.192.294,54	1.695.273,21
2010	1.397.210,08	1.980.667,65
2011	1.535.367,03	2.139.827,56
2012	1.782.498,88	2.434.281,10
2013	2.111.909,22	2.887.138,78
2014	2.247.378,95	3.053.584,53

Fonte - Elaborado a partir dos dados organizados por [17]

Primeiro passo é perceber por meio do gráfico de dispersão se existe indícios de uma relação de fato linear entre as variáveis.

Gráfico 11 - Diagrama de dispersão: Prestação de serviços (X) e PIB (Y)



Fonte: Construção própria.

Perceba que este diferente dos anteriores mostra um alinhamento perfeito entre os valores da amostra. O que caracteriza de fato uma relação de dependência linear entre elas, para verificar com veracidade iremos calcular o coeficiente de correlação que mensura o grau de associação entre as variáveis.

$$R_{xy} = \frac{7(2,68851.10^{13}) - (11297700, 30.15627239, 53)}{\sqrt{7(194.823.10^{13} - 1,27638.10^{14}}\sqrt{7(3,71121.10^{1}3 - 2,45149.10^{14})}}$$

$$R_{xy} = 0,99$$

Observe que de fato existe uma relação linear praticamente perfeita, ou seja, X terá relação direta com Y. Lembrando que isso irá ocasionar alterações significativas para esta situação em específico.

Podemos então encontrar os valores para α e β que melhor se ajustam representando os valores na forma de equação.

$$\beta = \frac{2,68851.10^{13} - 7(2139872,56.1535367,03)}{1,94823.10^{13} - 7(235735.10^{12})}$$

$$\beta = 1,2937$$

e α,

$$\alpha = 2.139.827,56 - (1,2937)1.353.367,03$$

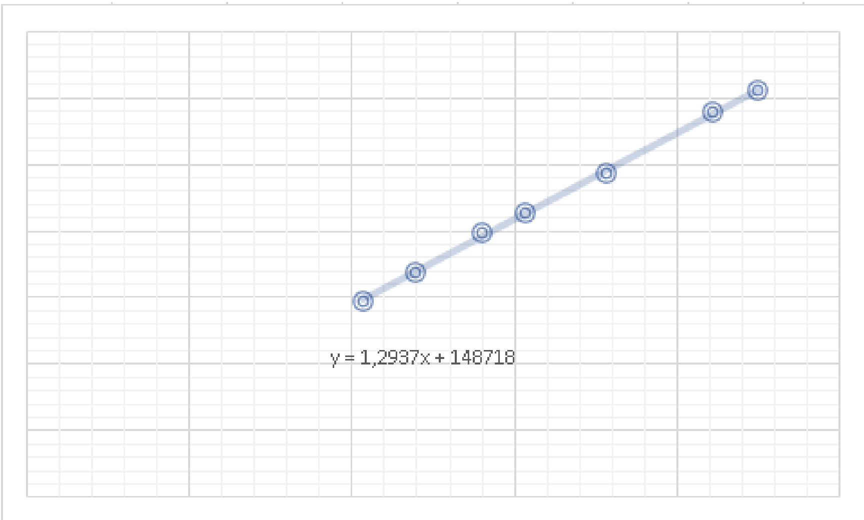
$$\alpha = 148718$$

Assim,

$$Y = 1,2937X - 148718$$

Por 4.4 temos que o coeficiente de determinação é $R^2 = 0,99$. Assim, há uma qualidade no ajustamento feito, e a equação reta explica bem os valores estudados. Graficamente, temos:

Gráfico 12 - Diagrama de dispersão: Prestação de serviços (X) e PIB (Y)



Fonte: Construção própria.

Iremos realizar alguns estudos a respeito dessa equação. Iremos considerar os seguintes valores para análise: $X_1 = 1.000.000,60$, $X_2 = 1.200.000,00$ e $X_3 = 2.000.000,00$.

Para X_1, temos:

$$Y_1 = 1,2937(1.000.000,60) + 148718$$

$$Y_1 = 1.144.982$$

Esse valor de X corresponde ao intervalos do período abaixo de 2008 e consequentemente seu valor correspondente para Y também ficou abaixo do valor do PIB no anos de 2008.

Para X_2, temos:

$$Y_2 = 1,2937(1.200.000,00) + 148718$$

$$Y_2 = 2.438.682$$

O valor de X nesse caso, está entre o período de 2009 à 2010, e calculamos então seu valor correspondente também dentro desse período.

Para X_3, temos:

$$Y_3 = 1,2937(2.000.000,00) + 148718$$

$$Y_3 = 3.732.382$$

Observe em 5.5 que os valores tidos acima estão no período de 2012-2013, e como percebemos a relação de dependência entre os valores de X e Y, temos que o valor de Y está exatamente dentro deste período.

Tabela 5.6: Valores em real por setor de atividade no PIB Araguaína-TO

X_i	Prestação de serviços(1000)	PIB (1000)
X_1	1.000.000,00	1.144.982
X_2	1.200.000,00	2.438.682
X_3	2.000.000,00	3.732.382

Fonte - Produção própria

Sendo assim, por essa relação de dependência, os valores alterados no setor de serviços acarretará mudanças no cálculo final do PIB nessa região de município. O que nos denota uma movimentação significativa no Setor de prestação de serviços indicando talvez a existência de grandes empresas e um número significativo de contratações para tal mudança nos valores do setor e a dependência dessas atividades no indicador econômico do município.

5.2 Estudo 2 - Análise de regressão linear múltipla

Concentraremos a análise do estudo do modelo de regressão linear múltipla em duas situações:

- PIB Y em função de Agropecuário X_1 e Prestação de serviços X_2;
- PIB Y em função de Industria X_1 e Prestação de serviços X_2.

Iremos poder meio de 4.19, 4.18 e 4.20 calcular o coeficientes que que melhor representam os valores da amostra em questão. Em seguida como feito na seção anterior, admitiremos um estudo para sabermos a interferência quando os valores de X_1 e X_2 variam.

5.2.1 Análise de regressão linear múltipla - Agropecuário, Prestação de serviços e PIB

Inicialmente manteremos ordem, calculando os coeficiente observando as variáveis da amostra em questão conforme a tabela 5.7.

Tabela 5.7: Valores em real por setor de atividade no PIB Araguaína-TO

Ano	Agropecuário (1000)	Prestação de serviços (1000)	PIB (1000)
2008	23.944,99	1.031.041,60	1.466.466,70
2009	30.466,67	1.192.294,54	1.695.273,21
2010	35.033,54	1.397.210,08	1.980.667,65
2011	35.719,61	1.535.367,03	2.139.827,56
2012	37.497,78	1.782.498,88	2.434.281,10
2013	37.313,83	2.111.909,22	2.887.138,78
2014	46.971,01	2.247.378,95	3.053.584,53

Fonte - Elaborado a partir dos dados organizados por [17]

Por 4.19, temos:

$$\hat{\beta}_1 = \frac{(297639301,7)(1,61501.1012) - (17447704294)(22668462743)}{(297639301,7)(11,24833.1012) - (13,04422.1020)}$$

$$\hat{\beta}_1 = 1,781089$$

Agora por 4.18, calcularemos β_1:

$$\hat{\beta}_2 = \frac{(22668462743) - 1,781089(17447704294)}{297639301,7}$$

$$\hat{\beta}_2 = 1,26884$$

E por fim, calcularemos β usando 4.20:

$$\hat{\beta} = 2.139.827,56 - 1,781089(35.278,20) - 1,26884(1.535.367,03)$$

$$\hat{\beta} = 126061$$

Dessa forma, podemos organizar a equação da reta de regressão 4.21 que melhor representa os valores de 5.7.

Assim,

$$\hat{Y} = 126061 + 1,781089.X_1 + 1,26884.X_2$$

Sendo assim, para quaisquer valores de X_1 e X_2 teremos um valor corresponde em Y. Dessa forma, iremos utilizar os valores já tabelados acima em 5.4 e 5.6 para supor valores de Y.

Assim iremos supor simuntaneamente os seguintes valores: ($X = 15.000,00$, $X = 22.900,00$ e $X = 26.900,99$) para X_1 e para para X_2 os valores ($X = 1.000.000,60$, $X = 1.200.000,00$ e $X = 2.000.000,00$).

Quando $X_1 = 15.000,00$ e $X_2 = 1.000.000,60$, teremos Y correspondente:

$$\hat{Y} = 126061 + 1,781089(15.000,00) + 1,26884(1.000.000,60)$$

$$\hat{Y} = 1.926.183,398$$

Quando $X_1 = 22.900,00$ e $X_2 = 1.200.000,00$, teremos:

$$\hat{Y} = 126061 + 1,781089(22.900,00) + 1,26884(1.200.000,00)$$

$$\hat{Y} = 1.689.455,631$$

E por fim, aos valores de $X_1 = 26.900,99$ e $X_2 = 2.000.000,00$ teremos:

$$\hat{Y} = 126061 + 1,781089(26.900,99) + 1,26884(2.000.000,00)$$

$$\hat{Y} = 2.711.653,548$$

Note que os valore de X_1 e X_2 aumentam, e o PIB também aumenta conforme os valores nos setores Agropecuário e Prestação de serviços aumenta, o que denota uma relação direta entre os valores e sua correspondência em Y. Conseguimos perceber claramente aqui, a relação de dependência entre as três variáveis em questão.

Assim, é possível dizer que quanto maior os valores de X_1 e X_2 maior será o valore d eY, e quanto menor estes valores menor será o calculo final do PIB.

5.2.2 Análise de regressão linear múltipla - Indústria, Prestação de serviços e PIB

Analisemos a seguinte tabela:

Tabela 5.8: Valores em real por setor de atividade no PIB Araguaína-TO

Ano	Indústria (1000)	Prestação de serviços (1000)	PIB (1000)
2008	215.509,05	1.031.041,60	1.466.466,70
2009	274.152,29	1.192.294,54	1.695.273,21
2010	309.754,84	1.397.210,08	1.980.667,65
2011	309.385,26	1.535.367,03	2.139.827,56
2012	333.638,52	1.782.498,88	2.434.281,10
2013	405.125,48	2.111.909,22	2.887.138,78
2014	402.454,77	2.247.378,95	3.053.584,53

Fonte - Elaborado a partir dos dados organizados por [17]

Por 4.18, 4.19 e 4.20 iremos calcular os coeficientes, como já feito da secção anterior.

$$\hat{\beta}_1 = \frac{(1,24833.10^{12})(2,34376.10^{11}) - (11,80212.10^{11})(1,61501.10^{12})}{(1,24833.10^{12})(27454809306) - (3,24764.10^{22})}$$

$$\hat{\beta}_1 = \frac{(2,92579.10^{23}) - (2,91044.1023)}{(3,42727.10^{22}) - (3,24764.10^{22})}$$

$$\hat{\beta}_1 = 0,854468696$$

Em $\hat{\beta}_2$,

$$\hat{\beta}_2 = \frac{(1,24833.10^{12}) - 0,854468696(11,80212.10^{11})}{1,24833.10^{12}}$$

$$\hat{\beta}_2 = 1,170380492$$

E por fim,

$$\hat{\beta} = 2.139.827,56 - 0,854468696(1.535.367,03) -,170380492(309.754,84)$$

$$\hat{\beta} = 73151,38027$$

Logo, a equação que representa os valores da amostra pelo modelo de regressão linear múltipla é:

$$\hat{Y} = 73151,38027 + 0,854468696X_1 + 1,170380492X_2$$

Assim para valores em $X - 1$ e X_2, teremos valores correspondentes em Y. Dessa forma, iremos usar os valores já usado nas análises de regressão linear simples para fazermos uma leitura aqui na regressão múltipla em que as variáveis regressoras são múltiplas.

Assim iremos supor o que acontece em Y quando é atribuído a equação os seguintes valores: $X_1 = 1.000.000,60$, $X_1 = 1.200.000,00$, $X_1 = 2.000.000,00$, $X_2 = 200.000,00$, $X_2 = 300.000,00$ e $X_2 = 400.000,00$.

Agora, analisando os dados atribuindo valores as variáveis regressoras da equação temos:

$$\hat{Y} = 73151,38027 + 0,854468696(1.000.000,60) + 1,170380492(200.000,00)$$

$$\hat{Y} = 1.161.696,69$$

Observe que o valor relacionado ao PIB, está abaixo dos ano de 2008 considerando que seus valores de X_i também estão abaixo deste ano. O que denota uma relação direta desses valores, é uma interferência direta das variáveis regressoras na variável Y.

$$\hat{Y} = 73151,38027 + 0,854468696(1.200.000,00) + 1,170380492(300.000,00)$$

$$\hat{Y} = 1.449.627,96$$

Nesse caso, os valores de X_i aumentara porém ainda ficam abaixo dos valores no ano de 2010, o que acarreta no valor do PIB também abaixo do ano de 2010 mais precisamente entre 2008 e 2009. Mostrando ainda interferência no cálculo final do PIB, considerando que há outra variáveis mais nesta soma final.

$$\hat{Y} = 73151,38027 + 0,854468696(2.000.000,00) + 1,170380492(400.000,00)$$

$$\hat{Y} = 2.250.240,97$$

Para este último caso, o valor do PIB se encontra abaixo do ano de 2013, o que por consequência nos diz que os valores de X_i também se mantêm abaixo desse período. Assim conseguimos observar que todo valor atribuído a X ocasionará em mudanças nos valores de Y, pois estão ligados a maior parte da produção do PIB, se não fosse isso, seus valores não teriam tanta interferência como vimos nessa análise. Dessa forma, de fato os setores Indústria e Prestação de serviços são setores com fortes contribuições para o PIB do município como já dito nesse trabalho.

Capítulo 6

CONSIDERAÇÕES FINAIS

O estudo da regressão no contexto financeiro é muito comum, por esta ser uma ferramenta de alta qualidade em análises de n variáveis quantitativas, o que nos possibilita um estudo quali-quanti. Esse trabalho teve o intuito de contribuir para algum eventual planejamento financeiro do município, ou para quem tenha interesse em obter conhecimentos acerca das informações do PIB dessa determinada região.

Iniciamos essa pesquisa questionando que estudos de regressão podem ser feitos com os dados do PIB de Araguaína-TO para servir como auxílio a qualquer planejamento econômico da cidade? Com o delinear da mesma, foi possível demonstrar por meio de cálculos matemáticos como os principais setores de fomento econômico de Araguaína-TO influenciam no cálculo final do PIB, usando os valores disponibilizados no período de 2008-2014.

No estudo de regressão linear simples foi possível perceber três situações e como estas em particular mantém relação com o PIB. Na primeira situação, onde vimos o poder do setor de Agropecuário no valor final do PIB, conseguimos perceber que é forte sua interferência e que alterações feitas para mais terão consequências positivas e aumento no PIB, se para menos terá influência negativa com uma queda nessa soma.

Na análise feita com o setor de Indústria e PIB, vimos que a influência do valor arrecadado pelo setor de Indústria não é tão forte, ou seja, não interfere de forma significativa na soma final assim, se sofre baixas não acarretará em reduções significativas no PIB. De igual modo, estudamos o setor de Prestação de Serviços e sua contribuição no PIB, onde concluímos que também é notável sua influência nesse cálculo final de produções no município, assim, suas mudanças ocasionarão mudanças no PIB.

Outro estudo foi movimentado nessa pesquisa, onde associamos duas variáveis para encontrar suas influências na soma final do PIB no estudo de regressão múltipla, assim, duas situações forma estudadas. A primeira delas, procuramos entender a força dos setores Agropecuário e Prestação de serviços no PIB, onde notamos grande prestígio, nesse ínterim percebemos que quando os valores desses setores sofrem alterações para mais ou para menos também o PIB sofrerá, denotando uma influência considerável.

E por fim, no estudo onde associamos Indústria e Prestação de serviços, onde notamos que é significativa a influência dessa relação no cálculo do PIB. Nesse estudo podemos fazer previsões à valores futuros dos PIB, caso quiséssemos aumentá-lo num certo ano, já saberíamos em que setores investir e quanto de aumento deveríamos ter nesse setor para que o PIB fosse elevado.

Assim, é evidente que a regressão linear tem ferramentas suficientes para estudos como estes. Aqui, conseguimos configurar a relação entres as variáveis, onde percebemos o grau de interdependência entre elas e sua influência na situação de interesse que é o PIB. Tendo em vista os resultados desta pesquisa, vemos que este estudo pode contribuir para eventuais planejamentos econômicos do município ou auxiliar em algum estudo.

Referências

[1] GUJARATY,D. N., et tal. **Econometria Básica.** Editora Afiliada. São Paulo, 2011.

[2] SOUSA, L. G. **Artigos de Economia Edição electrônica.** Disponível em: www.eumed.net/libros/2006b/lgs-art/. Acesso em: 12 maio 2021.

[3] MARTINS, Cristiana B. **O papel do dinheiro primitivo na economia Inca.** São Paulo: USP, 2001.

[4] SALMON, Wesley C. **Lógica.** 4. ed. Rio de Janeiro: Zahar, 1978.

[5] BÚRIGO, Fábio L.. **Moeda social e a circulação das riquezas na economia solidária(1).** fev. 2001. Disponível em: ¡https://lemate.paginas.ufsc.br/files/2016/06/Moeda-social-e-a-circulão-das-riquezas-na-economia-solidá1ria.pdf¿ Acesso em: 06 de junho de 2021.

[6] ECHEVESTE, Simone et al. **Educação estatística: perspectivas e desafios.** Acta Scientiae, Canoas, v. 7, n. 1, p. 103-109, jun. 2005. Disponível em: http://www.periodicos.ulbra.br/index.php/acta/article/viewFile/191/175. Acesso em: 02 nov. 2020.

[7] DEPARTAMENTO DE ESTATÍSTICA DA UFRN. **História da Estatística no Brasil e no Mundo.** 2016. Disponível em: https://estatisticaeetag.blogspot.com/. Acesso em: 07 dez. 2020.

[8] SAES, Flávio A. M.. SAES, Alexandre M.. **História da Economia Geral.** 1.ed. São Paulo: Saraiva, 2013.

[9] VERE-JONES, D. **The coming of Age Statistical Education.** International Statistical Review, Vol 63, Nº 1, Agosto 1995.

[10] SOUSA, Giselle Costa de; SCHIVANI ALVES, Juliana Maria. **A REGRESSÃO LINEAR DE GALTON: ATIVIDADES HISTÓRICAS PARA FUNÇÃO AFIM E ESTATÍSTICA BÁSICA USANDO PLANILHAS ELETRÔNICAS.** Conexões - Ciência e Tecnologia, [S.l.], v. 9, n. 4, p. 26-36, apr. 2016. ISSN 2176-0144. Disponível em: ¡http://conexoes.ifce.edu.br/index.php/conexoes/article/view/936/694¿. Acesso em: 16 may 2021.

[11] MLODINOW, L. **O andar do bêbado: como o acaso determina nossas vidas.** Rio de Janeiro: Zahar, 2009.

[12] MORETTIN, L. Gonzaga. **Estatística básica: probabilidade e estatística.** São Paulo: Pearson 2010.

[13] GONÇALVES, C. E. N.. **Introdução a economia.** Rio de Janeiro: Elsevier, 2010.

[14] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE) Disponível em ¡https://ibge.gov.br¿. Acessado em 29 de março de 2021.

[15] HOFFMAN, Rodolffo. **Análise de Regressão: uma introdução a econometria.** Piracicaba: ESALQqUSO,2015.

[16] PINHEIRO, João Ismael D.; DA CUNHA, Sônia Baptista; CARVAJAL, Santiago Ramírez; GOMES, Gastão Coelho. **Estatística Básica: a arte de trabalhar com dados.** Rio de Janeiro: Elsevier Editora, 2009.

[17] TOCANTINS. Secretaria do Planejamento e Orçamento. Diretoria de Pesquisa e Informações Econômicas. Perfil Socioeconômico dos Municípios. Palmas: SEPLAN-TO, 2017. 37 p.21.

[18] HELENE, Otaviano. **Método dos mínimos quadrados com formalismo matricial.** 1A.ED.2006.

[19] GIL, Antônio Carlos. **Como elaborar projetos de pesquisa.** 4. ed. - São Paulo :Atlas, 2002.