RICARDO DE SOUZA WILL

INTRODUÇÃO À ECONOMETRIA NO ENSINO MÉDIO: APLICAÇÕES DA REGRESSÃO LINEAR

Santo André, 2016

UNIVERSIDADE FEDERAL DO ABC

PROGRAMA DE MESTRADO PROFISSIONAL EM MATEMÁTICA

RICARDO DE SOUZA WILL

INTRODUÇÃO À ECONOMETRIA NO ENSINO MÉDIO: APLICAÇÕES DA REGRESSÃO LINEAR

Orientador: Prof. Dr. André Ricardo Oliveira da Fonseca

Coorientador: Prof. Dr. Antonio Cândido Faleiros

Dissertação de mestrado apresentada ao Programa de Mestrado Profissional em Matemática para obtenção do título de Mestre em Matemática.

ESTE EXEMPLAR CORRESPONDE A VERSÃO FINAL DA DISSERTAÇÃO
DEFENDIDA PELO ALUNO RICARDO DE SOUZA WILL,
E ORIENTADA PELO PROF. DR. ANDRÉ RICARDO OLIVEIRA DA FONSECA.

SANTO ANDRÉ, 2016

Sistema de Bibliotecas da Universidade Federal do ABC
Elaborada pelo Sistema de Geração de Ficha Catalográfica da UFABC
com os dados fornecidos pelo(a) autor(a).

Will, Ricardo de Souza
Introdução à Econometria no Ensino Médio : Aplicações da Regressão Linear / Ricardo de Souza Will. — 2016.

135 fls. : il.

Orientador: André Ricardo Oliveira da Fonseca
Coorientador: Antonio Cândido Faleiros

Dissertação (Mestrado) — Universidade Federal do ABC, Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, Santo André, 2016.

1. Noções de Estatística. 2. Conceitos de Econometria. 3. Regressão Linear Simples. 4. Regressão Linear Múltipla. I. Fonseca, André Ricardo Oliveira da. II. Faleiros, Antonio Cândido. III. Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, 2016. IV. Título.

MINISTÉRIO DA EDUCAÇÃO
Fundação Universidade Federal do ABC
Programa de Pós-Graduação em Mestrado Profissional em Matemática em Rede Nacional
Avenida dos Estados, 5001 – Bairro Santa Terezinha – Santo André – SP
CEP 09210-580 · Fone: (11) 4996-0017
profmat@ufabc.edu.br

FOLHA DE ASSINATURAS

Assinaturas dos membros da Banca Examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Ricardo de Souza Will, realizada em 14 de dezembro de 2016:

Prof.(a) Dr.(a) **Antonio Cândido Faleiros** (UFABC) – Presidente

Prof.(a) Dr.(a) **Ana Carolina Boero** (UFABC) – Membro Titular

Prof.(a) Dr.(a) **Antonio Carlos Gracias** (FEI) – Membro Titular

Prof.(a) Dr.(a) **Rafael de Mattos Grisi** (UFABC) – Membro Suplente

Prof.(a) Dr.(a) **Birajara Soares Machado** (HIAE) – Membro Suplente

Este exemplar foi revisado e alterado em relação à versão original, de acordo com as observações levantadas pela banca no dia da defesa, sob responsabilidade única do autor e com a anuência de seu orientador.

Santo André, 06 **de** MARÇO **de 20**17**.**

Assinatura do autor: ______________________

Assinatura do orientador: ______________________

Dedico este trabalho a minha esposa, meus filhos, netos e nora que estiveram ao meu lado em todos os momentos me apoiando, incentivando e descontraindo.

AGRADECIMENTOS

Agradeço a todos meus familiares, em especial a minha esposa, pelo apoio, colaboração e incentivo constante em todos os momentos de dificuldades.

À Sociedade Brasileira de Matemática pela implantação do PROFMAT e a CAPES pelo auxílio financeiro permitindo maior dedicação ao mestrado.

À Universidade Federal do ABC por abrir as portas ao PROFMAT e contribuir no aperfeiçoamento dos professores.

À todo corpo docente que atuam no PROFMAT, cuja disposição e empenho proporcionaram uma gana de conhecimento a todos alunos.

Ao Prof. Dr. André Ricardo Oliveira da Fonseca, meu orientador, pelo apoio científico e incentivo que me ajudou nesta etapa com muita competência, dedicação e paciência.

Ao Prof. Dr. Antonio Cândido Faleiros, meu coorientador, que aceitou em colaborar e contribuir, no qual fiquei extremamente grato.

Aos meus colegas do curso que ingressaram comigo neste programa e que ao longo deste período fortalecemos o vínculo de amizade e colaboração.

Aos gestores do ITB Prof. Antonio Arantes Filho que autorizaram a realização das atividades com os alunos e ter concedido o espaço nas salas de aulas e nos laboratórios de informática, ao Prof. Leonardo R. Pfiszter que colaborou em todas as etapas do projeto e aos alunos que participaram.

Aos professores que formaram a banca examinadora que aceitaram prontamente em participar e colaborar na finalização da dissertação.

"A maior de todas as ignorâncias é rejeitar uma coisa sobre a qual você nada sabe."

(H. Jackson Brown)

RESUMO

Esta dissertação tem como objetivo propor e dar subsídios aos professores de Matemática do 3° ano do ensino médio sobre temas envolvendo Estatística e Economia, tendo como sugestão a Econometria, especificamente a Regressão Linear Simples e Múltipla, em virtude de possuir conceitos muito abrangentes e que permitirá ao aluno desenvolver condições de entender as diversas aplicações e ser capaz de reconhecer o fenômeno linear e utilizar a regressão para fazer previsões.

Abordaremos no capítulo 1 uma revisão dos conceitos de Estatística dando ênfase ao desvio padrão, intervalos de confiança e teste de hipóteses.

No capítulo 2 teremos os conceitos de Econometria, como: a origem da palavra, a análise econométrica de um modelo matemático, os objetivos e a metodologia econométrica. Dando uma atenção em especial a Keynes e seus postulados sobre propensão marginal ao consumo e a poupar. Também permitirá aos alunos utilizarem seus conhecimentos de obtenção e tabulação dos dados das variáveis observadas, a construção do gráfico de dispersão, o ajustamento de uma reta que passa pelos pontos, determinar os parâmetros e a equação da reta.

No capítulo 3 trataremos do Modelo de Regressão Linear Simples. Inicialmente damos uma atenção especial a Galton que deu origem ao conceito de correlação, então passaremos para os cálculos dos parâmetros, dos resíduos, da variância, do desvio padrão e dos coeficientes de correlação e determinação. Utilizaremos os conceitos de estatística que foram relembrados no capítulo 1.

No capítulo 4 trataremos do Modelo de Regressão Linear Múltipla, portanto, abordaremos as diferenças quando deparamos com duas ou mais variáveis explicativas. O professor poderá revisar os conceitos de matrizes.

Por fim no capítulo 5 teremos o plano de aula, a escolha do público alvo e da unidade escolar, calendário e cronograma das atividades com os alunos e os resultados obtidos.

Palavras-chave: econometria, regressão linear, gráfico de dispersão, correlação linear

ABSTRACT

This dissertation aims to propose and to give subsidies to Mathematics teachers of the 3rd year of high school on topics involving Statistics and Economics, with Econometrics as a suggestion, specifically the Simple Linear Regression and Multiple, due to having very broad concepts and allow the student to develop a condition to understand the various applications and be able to recognize the linear phenomenon and use regression to make predictions.

We discuss in chapter 1 a review of statistical concepts emphasizing the standard deviation, confidence intervals and hypothesis testing.

In chapter 2, we will have concepts of Econometrics as the word's origin, the econometric analysis of a mathematical model, econometrics' goals and methodology. With a particular attention to Keynes and his postulates of small propensity for consuming and saving. It will also allow students to use knowledge of collecting and observed variables data tabulating, the construction of the scatter plot, adjustment of a line that passes through the points and determination of parameters and equation of the line.

In chapter 3, we will deal of the Simple Linear Regression Model. Initially we give a special attention to Galton that gave rise to the concept of correlation, then move on to the calculations of the parameters of waste, variance, standard deviation and to correlation and determination's coefficients. We will use the statistical concepts that were recalled in Chapter 1.

In chapter 4, we will treat the Multiple Linear Regression Model, therefore, we will discuss the differences when faced with two or more explanatory variables. The teacher may review the concepts of matrices.

Finally, in Chapter 5, we have the lesson plan, target audience and the school unit choice, calendar and schedule of activities with the students and the results obtained.

Keywords: econometrics, linear regression, scatter plot, linear correlation

CONTEÚDO

LISTA DE FIGURAS

LISTA DE TABELAS

INTRODUÇÃO

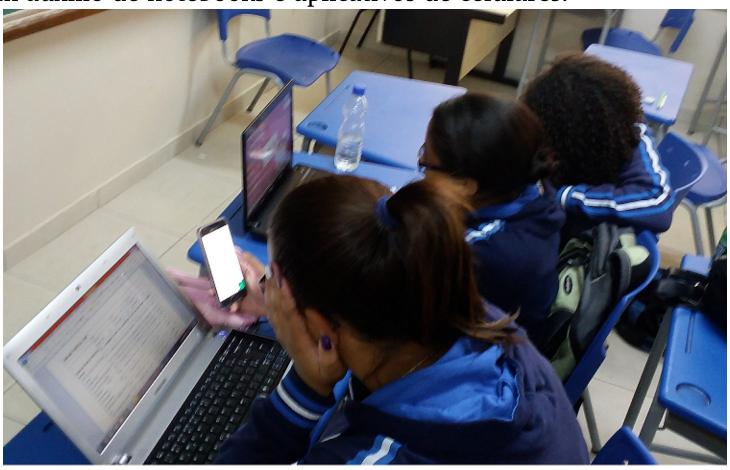
Atualmente está ocorrendo, em nosso país, discussões envolvendo uma série de mudanças na educação para a implementação do ensino médio integral com currículo flexibilizado, sendo que na proposta apenas três disciplinas permanecem como obrigatórias: matemática, português e inglês.

A proposta de reformulação do ciclo tem o objetivo de tornar o ensino médio mais atraente e diminuir os altos índices de evasão.

A Lei de Diretrizes e Bases da Educação (LDB) já defende a contextualização e as articulações das diversas disciplinas (a interdisciplinaridade e a transdisciplinaridade) que visam criar estratégias de integração disciplinar para reunir as possibilidades de conhecimento mais amplo propiciando a busca de soluções dos problemas e em oposição ao conhecimento monodisciplinar e desfragmentado.

A matemática na busca de atrativos para os alunos do ensino médio está em constante modernização e cabe aos professores estarem se aperfeiçoando e trazendo inovações ao currículo escolar, portanto, este trabalho tem a finalidade dar sugestões e ampliar as possibilidades em tornar suas aulas mais atraentes e direcionadas ao mercado atual, como pode ser visto na figura 1.

Figura 1: Alunos do ITB Prof. Antonio Arantes Filho, em Barueri, resolvendo as atividades em sala de aula com auxílio de notebooks e aplicativos de celulares.



Uma proposta é incluir no planejamento anual do 3° ano do ensino médio de Matemática um projeto envolvendo Estatística e Economia, tendo como tema a Econometria, especificamente a Regressão Linear Simples e Múltipla, podendo trabalhar interdisciplinarmente, como por exemplo, nas disciplinas: de Física, Geografia, História, Língua Portuguesa e Sociologia.

Este projeto tem o objetivo de permitir ao aluno desenvolver condições de entender as diversas aplicações e ser capaz de reconhecer o fenômeno linear e utilizar a regressão para fazer previsões e desenvolver conhecimento e habilidades importantes para a realização de análise econométrica, permitindo confrontar as teorias com os dados sócio-econômicos.

Conversando com colegas professores da Rede Estadual de São Paulo (região de Barueri, Carapicuíba, Jandira, Cotia e Itapevi) e da Rede Municipal de Barueri, a proposta foi bem recebida e nenhum colega se recorda de um professor ter trabalhado com os alunos este tema no ensino médio, no entanto, uma dificuldade de imediato é a falta de autonomia dos professores, pois o planejamento já vem pronto e engessado, inclusive os alunos das escolas estaduais já recebem o *Caderno do Aluno* com todo o conteúdo e exercícios propostos e inclusive são avaliados pelo menos quatro vezes ao ano através de uma Avaliação Diagnóstica de Matemática e Português, que são aplicadas no final de cada bimestre cobrando exclusivamente o conteúdo dos cadernos que é elaborada pela Secretaria de Educação Estadual e no caso do 3° ano do ensino médio ainda tem o SARESP, no entanto, a proposta foi bem aceita pelos professores e coordenação pedagógica, mas o tempo de duração do projeto e a preparação dos professores deveriam ser discutidos nos ATPCs (Aula de Trabalho Pedagógico Coletivo), que é desenvolvida na unidade escolar pelos professores e os coordenadores pedagógicos, que tem a finalidade de estimular o desenvolvimento das atividades coletivas da unidade escolar e articular os diversos segmentos da escola para a construção e implementação do seu trabalho pedagógico como por exemplo (re)planejar e avaliar as atividades de sala de aula.

A expectativa é incentivar e motivar os professores a saírem da zona de conforto e trabalhar projetos que motivem os alunos com temas do mundo atual.

1

NOÇÕES DE ESTATÍSTICA

Este capítulo tem a finalidade de revisar e reforçar alguns conceitos de estatística que foram identificados com a análise do questionário (figura 32 - Apêndice A) e solicitado na Atividade 1 do capítulo 5 e que serão abordados no trabalho.

Foi solicitado aos alunos a leitura e refazer os exemplos e os exercícios resolvidos e a assistir as vídeos aulas indicadas.

1.1 CONCEITOS

1.2 POPULAÇÃO E AMOSTRA

População é o conjunto de todos os elementos ou resultados cujas propriedades desejamos averiguar.

Muitas vezes, por motivos práticos ou econômicos, limitam-se os estudos estatísticos somente a uma parte da população, a **amostra**.

A amostra é um subconjunto finito de uma população que contém os elementos que podem ser observados e é onde as quantidades de interesse podem ser medidas.

1.3 PARÂMETROS E ESTIMADORES

Um **parâmetro** é representado geralmente por uma letra grega. Em Estatística é usado para designar uma grandeza mensurável, que caracteriza uma **população**. Na prática é comum que os parâmetros de uma população sejam desconhecidos, neste caso, os valores dos parâmetros são determinados através de um processo conhecido como **estimação**.

Um **estimador** é um valor característico de uma amostra, de tamanho n, extraída de uma população, de tamanho N, sobre a qual deseja-se tirar alguma conclusão. Os estimadores são representados por letras latinas.

Os principais estimadores são:

- A média da amostra, $\bar{x}$ que é um estimador da média da população: μ.
- A variância amostral, s^2 que é um estimador da variância populacional: σ^2.

A principal finalidade da estimação de parâmetros é o estudo de uma população com base em valores obtidos a partir de uma amostra extraída da mesma.

Podemos resumir com o auxílio de uma tabela 1 os parâmetros e os estimadores .

Tabela 1: Resumo sobre parâmetros e os estimadores

	Parâmetros da população	Estimadores pontuais
Média	μ	$\bar{x}$
Variância	σ^2	s^2
Desvio padrão	σ	s
Tamanho da população	N	n

1.4 VARIÂNCIA E DESVIO PADRÃO

A medida que dá o grau de dispersão (ou de concentração) de probabilidade em torno da média é a **variância** [25].

São as medidas de dispersão mais empregadas, pois levam em consideração a totalidade dos valores da variável em estudo. É um indicador de variabilidade bastante estável.

Para medir a dispersão dos dados em torno da média, os estatísticos usam a soma dos quadrados dos desvios dividida pelo tamanho da população. Definindo assim, variância como média aritmética dos quadrados dos desvios.

Observações [10]:

- Se os valores dos dados se repetirem em todas as amostras, então a variância da amostra será zero.
- Se os dados estiverem muito espalhados, então a variância da amostra acusará um número positivo elevado. Assim, uma grande variância significará uma grande dispersão dos dados em relação à média.
- A variância é uma medida que tem pouca utilidade na estatística descritiva, porém é extremamente importante na inferência estatística e em combinações de amostras.
- Quanto menor o desvio padrão, mais os valores da variável se aproximam de sua média.
- Quanto maior o desvio padrão, mais significativo à heterogeneidade entre os elementos de um conjunto

O **desvio padrão** é a medida de dispersão mais empregada, pois leva em consideração a totalidade dos valores da variável em estudo. É um indicador de variabilidade bastante estável. O desvio padrão baseia-se nos desvios em torno da média aritmética e a sua fórmula básica pode ser traduzida como: a raiz quadrada da média aritmética dos quadrados dos desvios e é representada por "s". Quanto menor o desvio padrão, mais parecidos são os valores da série estatística.

O desvio padrão indica a dispersão dos dados dentro da amostra, isto é, o quanto os dados em geral diferem da média.

Podemos calcular a variância e o desvio padrão utilizando:

	Amostra	População
Variância	$s^2 = \dfrac{\sum x^2 - n\bar{x}^2}{n-1}$	$\sigma^2 = \dfrac{\sum x^2}{n} - \mu^2$
Desvio padrão	$s = \sqrt{s^2}$	$\sigma = \sqrt{\sigma^2}$

1.5 NORMAL OU GAUSSIANA

Pelo **Teorema do Limite Central** (TLC) em uma **distribuição normal** uma **amostra aleatória simples** $(X_1, \cdots, X_n)$ é retirada de uma população com média μ e variância σ^2 . Então, temos que

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0, 1) \tag{1.1}$$

O TLC garante que para n grande a distribuição da média amostral, devidamente padronizada, se comporta segundo um modelo Normal padronizado (Z).

Em casos onde a verdadeira distribuição dos dados é simétrica, boas aproximações são obtidas para n ao redor de 30.

A distribuição normal pode ser considerada como a mais importante distribuição de probabilidade, pode ser aplicada em vários fenômenos. Também é conhecida como distribuição de Gauss.

A função de densidade da probabilidade normal pode ser calculada com

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)}, -\infty < x < \infty \tag{1.2}$$

No entanto vamos utilizar as tabelas de distribuição normal Z e t, que estão no Anexo.

1.6 INTERVALOS DE CONFIANÇA

A estimação de um parâmetro conduz a uma estimativa que é um número único, no entanto as amostras, frequentemente, não são muito grandes e, como consequência, os resultados obtidos com base nela são incertos [35].

Seja $(X_1, \cdots, X_n)$ uma amostra aleatória de uma população com a característica $X \sim f(x, \theta)$. Seja $T_1 = G(X_1, \cdots, X_n)$ e $T_1 = G(X_2, \cdots, X_n)$ duas estatísticas tais que $T_1 < T_2$ e que

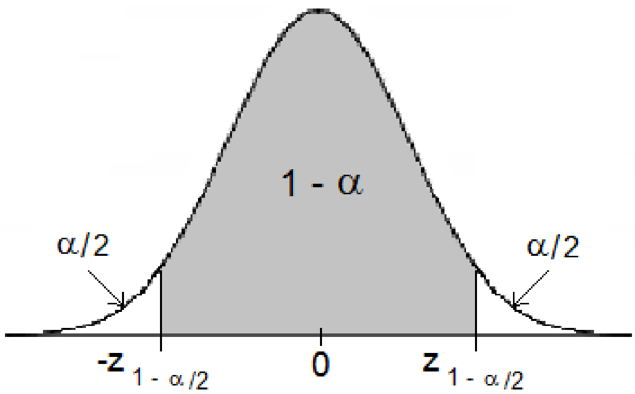
$$P(T_1 < \theta < T_2) = 1 - \alpha \tag{1.3}$$

O intervalo (T_1, T_2) é chamado de intervalo de confiança para θ, onde T_1 e T_2 são os limites inferior e superior e respectivamente e $1-\alpha$ é o coeficiente (ou nível) de confiança.

Logo, fixando um nível de confiança $1-\alpha$, pode-se determinar $z_{\alpha/2}$ de tal forma:

$$P\left(-z_{1-\frac{\alpha}{2}} \leqslant Z \leqslant z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \tag{1.4}$$

Figura 2: Nível de confiança



$$(\mu, 1-\alpha) = \left(\bar{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) \tag{1.5}$$

Logo, o intervalo de confiança para μ é dado por

$$IC(\mu, 1-\alpha) = \left(\bar{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \quad ; \quad \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = (\bar{X} - E; \bar{X} + E) \tag{1.6}$$

Resumindo, devemos saber a origem do desvio padrão, pois será importante para escolher a tabela Z ou t de distribuição.

Uma ajuda para definir qual distribuição utilizar é dado pela tabela 2.

Tabela 2: Escolha apropriada da Tabela de Distribuição Z ou t.

Tamanho	σ da população	s da amostra
$n \geqslant 30$	$\bar{x} \pm Z \cdot \frac{\sigma}{\sqrt{n}}$ tabela Z	$\bar{x} \pm Z \cdot \frac{s}{\sqrt{n}}$ tabela Z
$n < 30$	$\bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$ tabela Z	$\bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$ tabela t

1.7 EXEMPLO

1-) Uma clínica pretende estimar o tempo médio gasto em cada consulta, foram sorteados 64 pacientes. Essa amostra indicou um tempo médio de atendimento de 10 minutos e um desvio padrão de 3 minutos. Com base nestes dados, determine o tempo médio de atendimento a um nível de confiança de 90% [13].

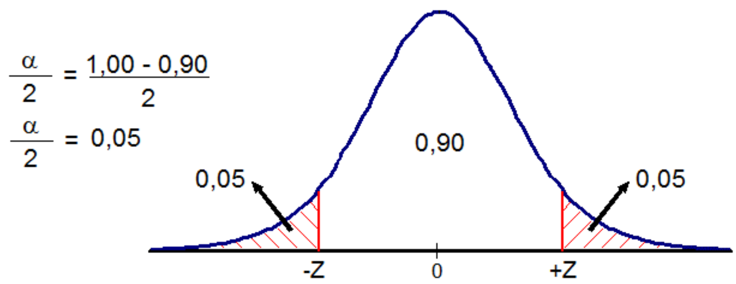
Resolução

Dados: $n = 64$, $\bar{x} = 10$ e $s = 3$

Neste caso, vamos utilizar a tabela Z do Anexo A (ver tabela 2).

Então, com base na tabela Z de escores negativos (Figura 34) teremos para a $\alpha/2 = 0,05$ (Figura 3) temos um $Z \pm 1,64$), pois a distribuição é simétrica.

Figura 3: Distribuição normal do tempo de espera dos pacientes



Então: $\bar{x} \pm Z \cdot \frac{s}{\sqrt{n}} = 10 \pm 1,64 \cdot \frac{3}{\sqrt{64}} = 10 \pm 0,62$

Portanto nosso erro máximo é de 0,62 com 90% de confiança com um intervalo de confiança de

$$IC \quad (10 - 0,62 \quad ; \quad 10 + 0,62)$$

$$IC \quad (9,38 \quad ; \quad 10,62) \quad minutos$$

1.8 TESTES DE HIPÓTESES

É de grande importância a verificação de hipóteses. Admite-se a hipótese em prova como verdadeira e calcula-se a probabilidade de que nossa amostra ocorra nessas condições. Se esta probabilidade é maior que o nível de significância fixado (por exemplo 1% ou em 5%) a hipótese será admitida como verdadeira, consequentemente se essa probabilidade for menor que o nível de significância, será pouco provável que a hipótese em prova ocorra, portanto devemos rejeitá-la [35].

1.9 EXERCÍCIOS

1.9.1 *Teste de Hipóteses*

1. O responsável pelo controle de qualidade de uma fábrica que vende suco de laranja, supervisiona periodicamente a linha de produção para verificar se as caixas de sucos de laranja estão com a quantidade correta de suco, isto é, verifica se não está ocorrendo sobreenchimento ou subenchimento da caixinha. Em média, o volume das caixas de suco é de 1,5 litros e o desvio padrão populacional é de 0,2 litros. Para testar a hipótese de que haja a necessidade de recalibrar a máquina que enche as caixas, foi retirada uma amostra de 35 caixas de suco. A média encontrada foi de 1,43 litros. Ao nível de significância de 5%, teste as hipóteses e diga sua conclusão [31].

   Resolução

   Dados: $n = 35$, $\mu = 1,5\ell$, $s = 0,2$, $\bar{X} = 1,43\ell$ e $\alpha = 0,05$

   Hipóteses:
   $H_0 : \mu = 1,5\ell$
   $H_1 : \mu \neq 1,5\ell$
   Com base nas hipótese teremos um teste bilateral.

Como o desvio padrão populacional é conhecido, então, a distribuição é normal, portanto teremos um Z crítico (Z_c).

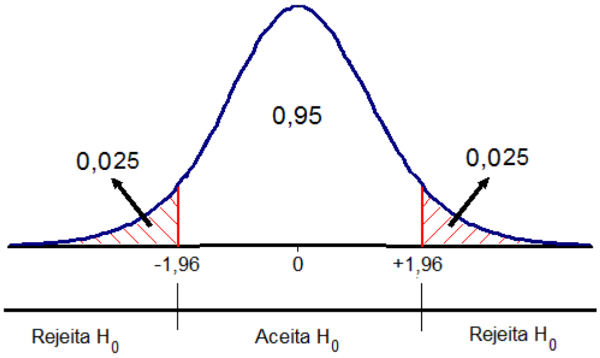
O nível de significância determina as áreas que temos nas extremidades. Então, como o teste é bilateral a área de cada extremidade é metade do nível de significância ($\alpha/2$), portanto é de 0,025.

Como estamos utilizando a distribuição normal para determinar o Z_c utilizaremos a tabela (Figura 34) do anexo.

Este valor corresponde a $1,96$, portanto teremos os valores: $Z_c = -1,96$ e $Z_c = +1,96$.

Portanto, verificaremos se as hipóteses serão rejeitadas ou aceitas. Temos a rejeição de H_0, conforme figura 4.

Figura 4: Teste de hipóteses.



Agora vamos calcular a estatística de teste.

$$Z_{Teste} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{1,43 - 1,5}{\frac{0,2}{\sqrt{35}}} = -2,07$$

O valor de Z_{Teste} caiu na área de rejeição, então devemos rejeitar a hipótese nula (H_0), portanto, a caixa não está sendo enchida devidamente e existe evidências suficientes para que o responsável pelo controle de qualidade pare a linha de produção e calibre a máquina.

2. A emissão média de gases poluentes de motores com uma nova tecnologia precisa ser menor que 20 ppm (partículas por milhão) para estar nas novas regulamentações de emissão de poluentes. Dez motores são produzidos para fazer os testes, e o nível de emissão de poluente de cada um é determinado. Os dados de emissão são:

$$15,6; \quad 16,2; \quad 22,5; \quad 20,5; \quad 16,4; \quad 19,4; \quad 16,6; \quad 17,9; \quad 12,7; \quad 13,9.$$

Os dados apresentados são evidências suficientes para concluir que esse tipo de motor está de acordo com os novos padrões? Considere um risco de erro com probabilidade de 1% [18].

Resolução

Fazendo uma tabela teremos:

Tabela 3: Emissão média de gases poluentes de motores

Motor	X_i	$(X_i - \bar{X})^2$
1	15,6	2,46
2	16,2	0,94
3	22,5	28,41
4	20,5	11,09
5	16,4	0,59
6	19,4	4,97
7	16,6	0,32
8	17,9	0,53
9	12,7	19,98
10	13,9	10,69
Soma	171,7	80,00
Média	17,17	

Fazendo a média amostral de nossos dados (n = 10) teremos $\bar{X} = 17,17$ e o desvio padrão amostral será calculado pela equação (3.19)

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{80}{9}} = 2,98$$

Hipóteses:
$H_0 : \mu = 20 \quad ppm$
$H_1 : \mu < 20 \quad ppm$

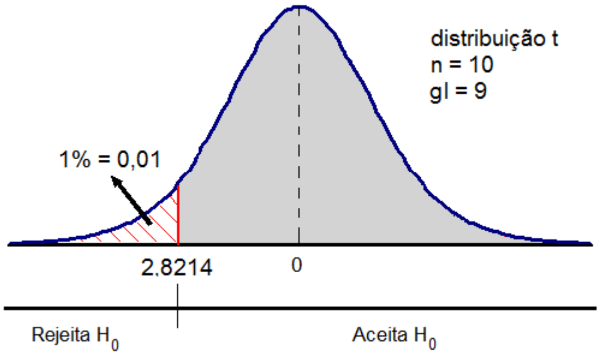
Devemos considerar H_0 como verdadeira e rejeitar H_0 se $P(\bar{X} = 17,6|H_0) < 1\%$

Que tipo de distribuição devemos utilizar, pois nossa amostra é muito pequena?

Neste caso, vamos utilizar uma estatística t (ver Tabela 2), então vamos calcular Z_{Teste}.

$$Z_{Teste} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{17,17 - 20}{\frac{2,98}{\sqrt{10}}} = -3,00$$

Figura 5: Teste de hipóteses e determinação do Z.



Podemos observar que é uma distribuição unicaudal (figura 5) e utilizando a tabela (Figura 33) do anexo, sendo que o grau de liberdade é $gl = n - 1 = 10 - 1 = 9$, e que a área seja de 1%. Neste caso encontramos o valor corresponde a $2,8214$, como a distribuição t é simétrica, nosso valor procurado será $-2,8214$, portanto, o valor de Z_{Teste} caiu na área de rejeição, pois $-3,00 < -2,8214$, então a probabilidade de ocorrer o evento é muito menor que 1%, em virtude disso devemos rejeitar a hipótese nula (H_0) e aceitar a a hipótese alternativa (H_1) e poder afirmar que este tipo de motor está de acordo com a nova regulamentação.

1.9.2 *Intervalo de Confiança*

1. Uma gravadora de músicas mediu o tempo de duração de 50 CDs. O tempo médio obtido foi de 61,8 minutos, com um desvio padrão de 3,5 minutos. Determine, com o nível de confiança de 95% [13]:
   a) O erro máximo do tempo médio.
   b) A estimativa do intervalo de confiança.

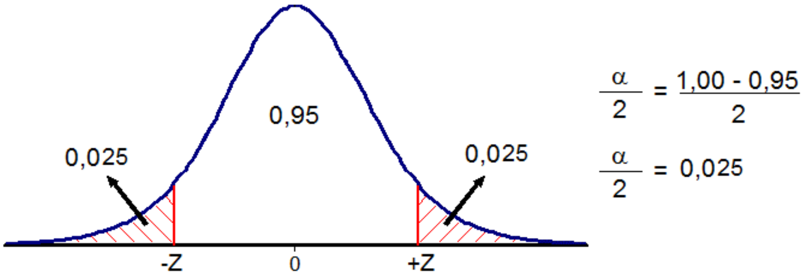
Resolução

Dados: $n = 50$, $\bar{x} = 61{,}8$ e $s = 3{,}5$ minutos

Neste caso, vamos utilizar a tabela Z (ver Tabela 2).

Então, com base na tabela Z de escores negativos (Figura 34) teremos para a $\alpha/2 = 0,025$ (Figura 6) temos um $Z \pm 1,96 \geq 4$), pois a distribuição é simétrica.

Figura 6: Distribuição normal do tempo de músicas dos CDs



Então: $\bar{x} \pm Z \cdot \dfrac{s}{\sqrt{n}} = 61,8 \pm 1,96 \cdot \dfrac{3,5}{\sqrt{50}} = 61,8 \pm 0,97$

Portanto, nosso erro máximo é de 0,97 minutos com 95% de confiança com um intervalo de confiança de no tempo de duração, em minutos, de cada CD.

$$IC(61,8 - 0,0,97 \quad ; \quad 61,8 + 0,97)$$

$$IC(60,83 \quad ; \quad 62,77) \quad minutos$$

2

CONCEITOS DE ECONOMETRIA

2.1 O QUE É ECONOMETRIA?

A palavra **econometria** vem de duas palavras de origem grega: *oikonomia*, que significa economia e *metron* que indica medida, medição [35].

O nome econometria foi introduzido em 1926, pelo economista e estatístico norueguês Ragnar Frisch [22]. Ele foi o pioneiro em desenvolver um modelo estocástico [1] do ciclo econômico [34].

Em termos literais, econometria significa "medição econômica". Consiste na aplicação da estatística matemática a dados econômicos para dar suporte empírico[2] aos modelos formulados pela economia matemática e obter dados numéricos a problemas de economia [14].

A econometria é uma combinação de três ramos de ciências: Economia, Matemática e Estatística, com o objetivo de dar conteúdo empírico às formulações teóricas da Economia [23].

A econometria vem depois da teoria. Um observador formula uma teoria, postulados, hipóteses e suposições ao observar a realidade e partir daí surge um modelo matemático.

Essa teoria é possível ser escrita matematicamente através de equações e o econometrista usa estas equações matemáticas formuladas geralmente pelo economista

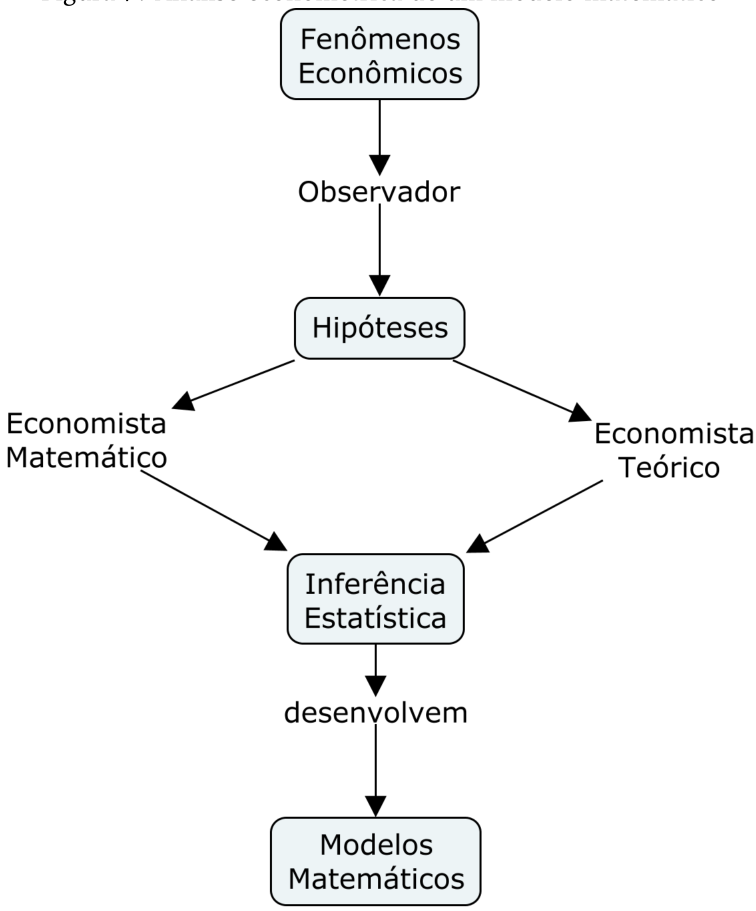
1 modelo estocástico ou matemático: cujas variáveis respondem a uma distribuição específica. Tais modelos não oferecem soluções únicas, mas apresentam uma distribuição de soluções associadas a uma probabilidade.

2 empírico é um fato que se apóia somente em experiências vividas, na observação de coisas, e não em teorias e métodos científicos.

matemático e se utiliza de dados da realidade e verifica que é um suporte empírico, porque ele objetiva dar sustentação a esta teoria com base na aplicação de dados reais a esta teoria.

A econometria pode ser definida como a análise quantitativa dos fenômenos econômicos, isto é, enquanto o observador está observando a realidade e formulando hipóteses, o economista matemático juntamente com o economista teórico com base no desenvolvimento paralelo da teoria e das observações e com o uso de métodos de inferência[3] adequados desenvolvem modelos matemáticos e juntamente desenvolve-se a econometria em cima das observações estamos falando de estatística aplicada em economia (Figura 2).

Figura 7: Análise econométrica de um modelo matemático



Na aplicação da econometria podemos ter várias dificuldades como por exemplo: o problema metodológico da especificação das variáveis independentes, a correta de-

3 Inferência é fazer afirmações a partir de um conjunto de valores representativo sobre um universo. Tal tipo de afirmação deve sempre vir acompanhada de uma medida de precisão sobre sua veracidade

finição da forma do modelo e a adequada utilização da estatística como método de análise [26].

A econometria é uma ciência de cunho social porque é a estatística aplicada a economia que estuda a interação entre os agentes econômicos, o comportamento de pessoas, de empresas, de famílias, do governo, então este comportamento a toda hora está mudando. Ele decorre da interação entre seres humanos, por este motivo não podemos falar que a econometria não é uma ciência exata, por mais que ela seja uma ciência quantitativa. A quantificação destes dados ela se dá como base o comportamento social dos agentes econômicos.

2.2 QUAIS OS OBJETIVOS DA ECONOMETRIA?

O econometrista ao fazer uma análise do problema econômico deverá escolher um modelo que seja capaz de satisfazer os objetivos de explicar os comportamentos das variáveis observadas ou prever comportamentos ainda não observados, ou ambos [5].

A Econometria tem por objetivo distinguir as seguintes fases [2] [14] [23]:

- Investigação das teorias econômicas;
- Formulação de teste e hipóteses sobre o comportamento da realidade;
- Utilização da matemática e da estatística para formular modelos matemáticos e econométricos;
- Mensuração[4] das variáveis econômicas;
- Estimação dos parâmetros de relações estabelecidas pela teoria econômica;
- Prever valores futuros de variáveis econômicas com o uso do modelo adequado.

4 mensuração é o ato de medir, ou seja, de determinar o valor de certas grandezas. É sinônimo de medição.

2.3 ANÁLISE ECONOMÉTRICA

A econometria é a aplicação dos métodos estatísticos aos problemas econômicos, portanto é uma técnica de estatística aplicada. O emprego adequado destas técnicas, que se baseiam geralmente em análises de regressão, depende de um entendimento dos métodos estatísticos envolvidos, mas a interpretação dos resultados também depende da habilidade para perceber a importância relativa do modelo teórico e dos dados empíricos [1].

De acordo com Gujarati: "... a econometria é uma amálgama[5] de teoria econômica, economia matemática, estatística econômica e estatística matemática [14]."

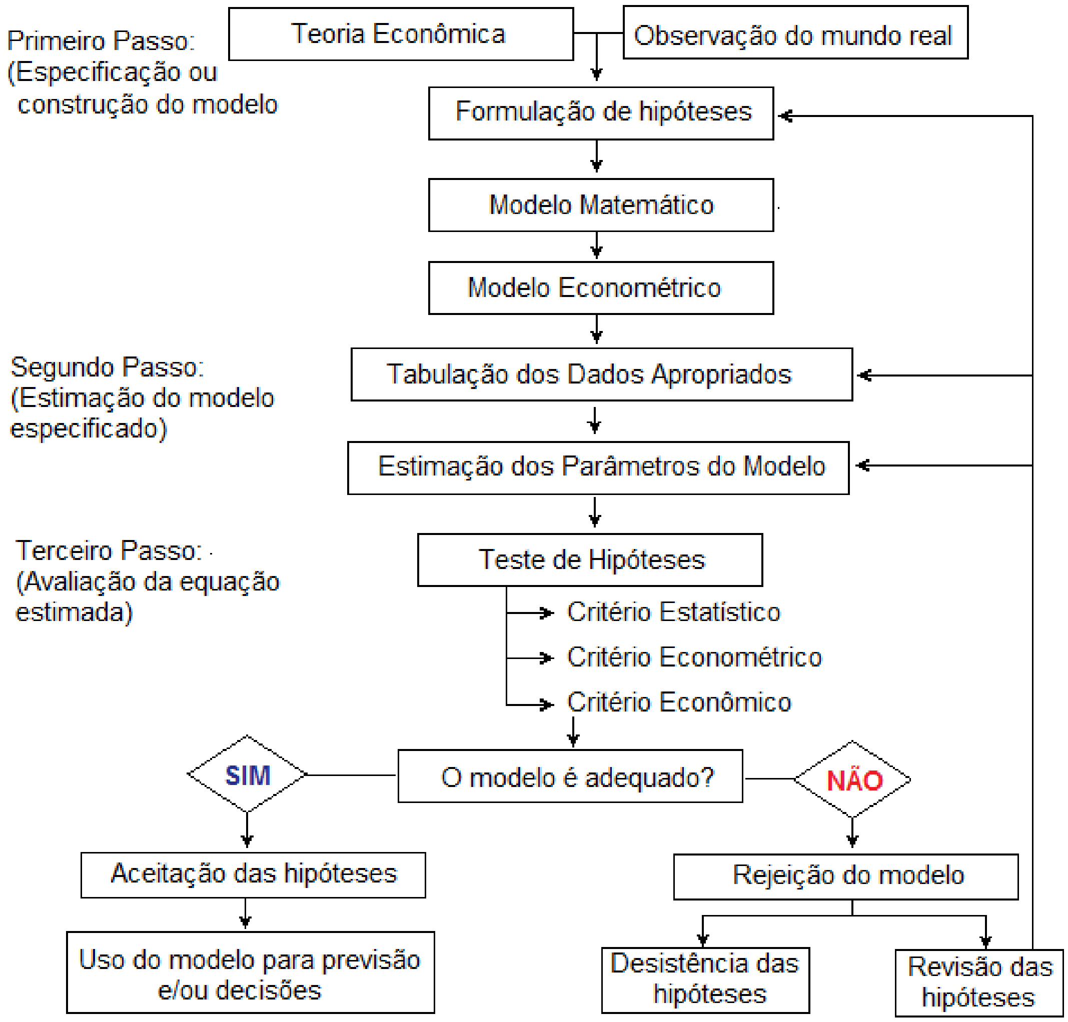
- A teoria econômica faz hipóteses da natureza qualitativa. Por exemplo a teoria do consumidor, um produto que tenha o preço aumentado faz com que a sua quantidade consumida diminua. Ela faz esta declaração apenas no caráter qualitativo e não analisa a magnitude e a quantidade que esta demanda vai cair, por isto é necessário que venha a economia matemática.
- A economia matemática formula inúmeras equações sem levar em conta se a teoria pode ser medida ou verificada. Então cabe ao econometrista fazer isto, mas precisamos da estatística econômica.
- A estatística econômica tem a função de buscar, coletar, processar e apresentar os dados econômicos. Um exemplo o economista estatístico do IBGE é responsável por levantar os dados do PIB, faz o levantamento dos dados, processa estes dados, apresenta estes dados, que na realidade são dados brutos, em forma de gráficos ou tabelas, enfim ele fica limitado apenas isto.
- A estatística matemática proporciona as ferramentas usadas na atividade do econometrista, como métodos especiais de natureza específica, assim como ocorre a inferência estatística. Ele aplica os dados que foram coletados ao modelo econométrico e faz os testes estatísticos para verificar se pode aceitar ou refutar o modelo.

5 Amálgama é também o nome que se dá à mistura de coisas diversas e heterogêneas.

2.4 METODOLOGIA ECONOMÉTRICA

A metodologia econométrica é um conjunto de métodos ou práticas que define se um modelo será aceito ou refutado. Ela tem algumas etapas que podem ser divididas em três passos que devem ser respeitada sequencialmente (Figura 8, adaptada de [23]).

Figura 8: Metodologia da investigação econométrica

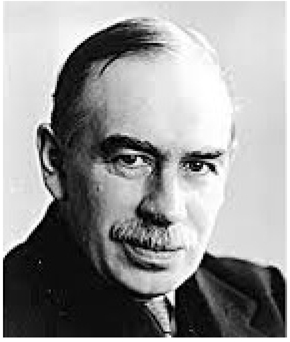


Vamos explicar cada uma das etapas utilizando exemplo da relação entre consumo e renda de keynesiana.

2.4.1 *Propensão Marginal ao Consumo e Propensão Marginal a Poupar*

John Maynard Keynes [6] (The General Theory of Employment, Interest, and Money, 1936) defende que o consumo dependeria primordialmente do nível de renda [33].

Figura 9: John Maynard Keynes



Keynes postulava que a **propensão marginal ao consumo** (PMC) é um valor que representa nível de aumento no consumo para cada unidade monetária aumentada na renda, sendo este variando entre 0 e 1, que expressa o comportamento dos agentes de uma determinada economia em relação a renda disponível (a renda disponível depois do pagamento dos impostos) [19].

A função de consumo (equação (2.1)) expressa a relação entre o nível de renda (X_t) e o nível de consumo no tempo (C_t). Tanto X e C se expressam em milhares de milhões de unidades monetárias [33].

$$C_t = f(X_t) \tag{2.1}$$

6 John Maynard Keynes (1883 - 1946) foi um economista britânico que revolucionou a teoria macroeconômica de tradição neoclássica. Na época, o mundo enfrentava um período de colapso do capitalismo, provocado pela quebra da bolsa de Nova York em 1929 – o acontecimento econômico mais emblemático do século XX. Em seus trabalhos, Keynes tinha como objetivo encontrar uma saída para salvar o capitalismo, modelo que não defendia fervorosamente, mas que acreditava ser a melhor garantia de evolução da sociedade, em um momento de surgimento de fortes correntes socialistas. Seu pensamento ficou marcado pela discussão do papel do estado na economia e dos mecanismos que poderiam ser usados para reativá-la nas condições de depressão, tais como os gastos públicos e a política fiscal [37].

Essa PMC revela qual a porcentagem, a comunidade designa ao consumo. Ela é definida como a derivada do consumo em relação a renda (equação (2.2)). Lembrando que $0 < PMC < 1$.

$$PMC = \frac{dC}{dX} \tag{2.2}$$

Mas apenas uma fração desta variação de renda disponível (dX) está destinada para a variação ao consumo (dC), pois uma parte da variação da renda fica destinada a poupar (dS)(equação (2.3)).

$$dS = dX - dC \tag{2.3}$$

A **propensão marginal a poupar (de poupança)** (PMS) refere-se à tendência para poupar mediante o rendimento disponível. Podemos derivar ambos os lados da equação (2.3) em relação a renda.

$$\frac{d}{dX}(S) = \frac{d}{dX}(X) - \frac{d}{dX}(C) \tag{2.4}$$

$$\frac{dS}{dX} = 1 - \frac{dC}{dX} \tag{2.5}$$

Portanto temos:

$$PMS = 1 - PMC \tag{2.6}$$

$$PMC + PMS = 1 \tag{2.7}$$

Portanto, temos na equação (2.7) que a todo aumento da renda disponível será dividido entre a PMC e PMS, conforme equação (2.3).

2.4.2 *Formulação de Hipóteses*

Como dito antes, isto não cabe ao econometrista, isto cabe ao economista teórico que baseado nas suas observações da realidade ele escreve a teoria, isto é, as hipóteses e suposições.

A econometria utiliza de dois ingredientes básicos: teoria e fatos. Temos que buscar e trabalhar com os dados sempre a partir de uma teoria. E uma das teorias mais utilizadas para iniciar o estudo da econometria é a teoria do consumo, pois vem sendo uma grande fonte de integração entre micro e macroeconomia moderna, então vamos utiliza-la para descrever a metodologia econométrica.

2.4.3 *Especificação do Modelo Matemático*

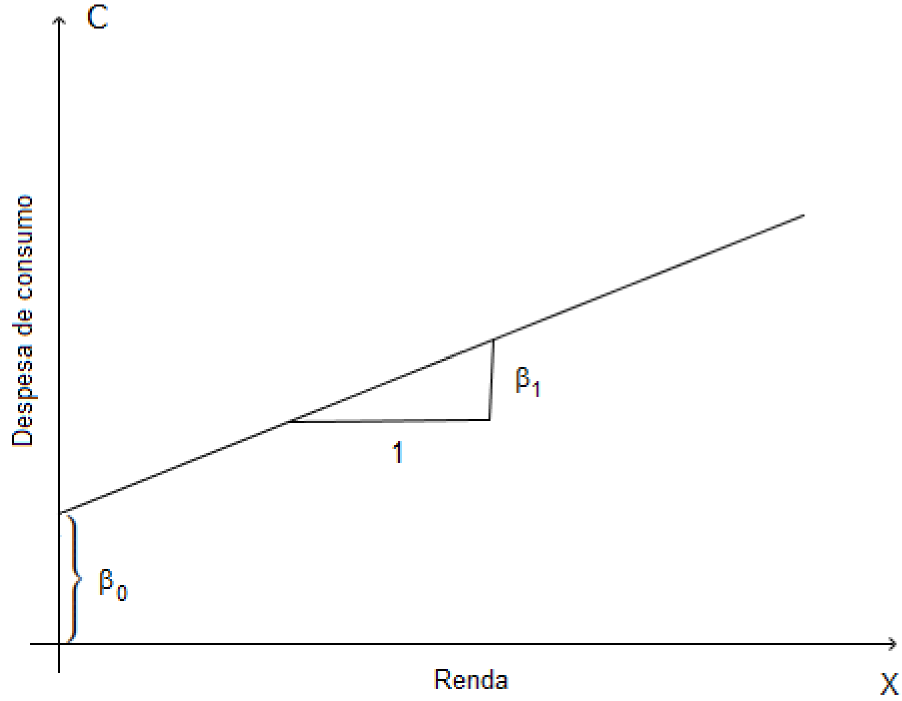
Cabe ao economista matemático com base nas hipóteses apresentadas escrever este modelo de forma matemática com uso de equações.

De acordo coma teoria de Keynes um economista matemático irá transformar a função de consumo (equação (2.1)) em uma equação de consumo keynesiana [14]:

$$C_t = \beta_0 + \beta_1 X_t, \quad onde \quad 0 < \beta_1 < 1 \tag{2.8}$$

Dando origem ao gráfico da função consumo keynesiana.

Figura 10: Gráfico da função consumo keynesiana.



Lembrando que C_t = despesas de consumo no tempo e X_t = renda disponível, $\boldsymbol{\beta_0}$ e $\boldsymbol{\beta_1}$ são os **parâmetros** da reta de regressão do modelo matemático, sendo $\boldsymbol{\beta_0}$ o **intercepto** e $\boldsymbol{\beta_1}$ o **coeficiente angular**. O intercepto é o valor da reta de regressão para X = 0; é o ponto em que a reta de regressão cruza o eixo C, como mostra a figura 10. Neste caso, mesmo que a renda seja zero deverá haver consumo para as necessidades básicas humanas que devem ser atendidas independente da renda, por exemplo: alimentação.

A equação (2.8) é o **modelo de regressão linear com um único regressor ou modelo uniequacional**, em que C é a **variável dependente** e X é a **variável independente ou regressor**. Essa equação relaciona linearmente o consumo à renda, é um exemplo de modelo matemático, onde um modelo nada mais é que "uma representação simplificada da realidade".

2.4.4 *Especificação do Modelo Econométrico*

O modelo matemático representado pela função consumo apresentado pela equação (2.8) é limitado para o econometrista, pois supõe que exista uma relação precisa entre o consumo e a renda, mas precisão é algo que não existe na econometria e na estatística, pois trabalhamos com o comportamento social, que não é exato, não é preciso, e muda constantemente, pois existem fatores que não são observados diretamente na equação matemática e que influenciam no comportamento da variável dependente. Estes fatores estão implícitos [7].

O modelo econométrico ou estatístico é um modelo que tem as especificações necessárias para sua aplicação empírica.

A hipótese é que a renda é o principal fator que sustenta a função consumo, mas não é o único, teremos outros fatores que foram omitidos e que se forem considerados provavelmente se encontrarão fora da reta de regressão e portanto teremos dispersão de pontos, portanto o econometrista deve modificar a equação consumo de forma mais geral como:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon \tag{2.9}$$

7 Implícitos: que se apresenta de modo obscuro; que está ou permanece subentendido; que não se pode expressar formalmente; não declarado; oculto, tácito.

O termo ε na equação (2.9) é o **termo de erro**, conhecido como distúrbio, é uma variável aleatória (estocástica) que tem propriedades probabilísticas e que afetam o consumo, mas que no modelo matemático não são incorporados em virtude da impossibilidade de medi-los ou de seu desconhecimento [14]. Entretanto, do ponto de vista dos modelos econométricos, esta especificação adquire uma grande relevância, portanto engloba aspectos que não devem ser desprezados em toda investigação empírica. Dentro desta ordem os modelos devem levar em conta tanto os erros de observação como as perturbações [2].

2.4.5 *Obtenção e tabulação dos dados*

É necessário obter os dados econômicos, dados estatísticos da economia do país, de uma região, de um Estado, de um município, de um bairro, de uma empresa, de um setor econômico, isto depende muito a que nível vamos utilizar (microeconomia ou macroeconomia).

Podem ser através de resultados das observações de experimentos não controlados que sejamos capazes de construir um modelo econômico que seja capaz de obter informações e proporcionar a percepção sobre os parâmetros econômicos desconhecidos. A maioria dos dados econômicos são coletados para fins administrativos e principalmente para agências governamentais. Estes dados podem ser coletados em forma de [15]:

- *séries temporais*: dados coletados ao longo de intervalos discretos de tempo, seja diariamente (preço de ações, relatórios meteorológicos), mensalmente (taxa de desemprego, IPC), trimestralmente (PIB).
- *cortes transversais no tempo*: dados coletados sobre unidades de amostra em determinado período de tempo (pesquisas de opinião, dados de censos).
- *painel de dados*: dados que acompanham microunidades individuais ao longo do tempo (PIB de cada país sul-americano para o período de 1990 a 2008).

No nosso exemplo vamos utilizar os dados de despesas de consumo pessoal e produto interno bruto do EUA no período 2000-2014, do Economic Report of the President, 2014, em bilhões de dólares (Tabela 4) [6].

Tabela 4: Despesas de Consumo Pessoal (C) e Produto Interno Bruto (X) dos EUA no período 2000-2014.

Ano	PIB (X)	DCP (C)
2000	10.284,8	6.792,4
2001	10.621,8	7.103,1
2002	10.977,5	7.384,1
2003	11.510,7	7.765,5
2004	12.274,9	8.260,0
2005	13.093,7	8.794,1
2006	13.855,9	9.304,0
2007	14.477,6	9.750,5
2008	14.718,6	10.013,6
2009	14.418,7	9.847,0
2010	14.964,4	10.202,2
2011	15.517,9	10.689,3
2012	16.163,2	11.083,1
2013	16.768,1	11.484,3
2014	17.420,7	11.928,4

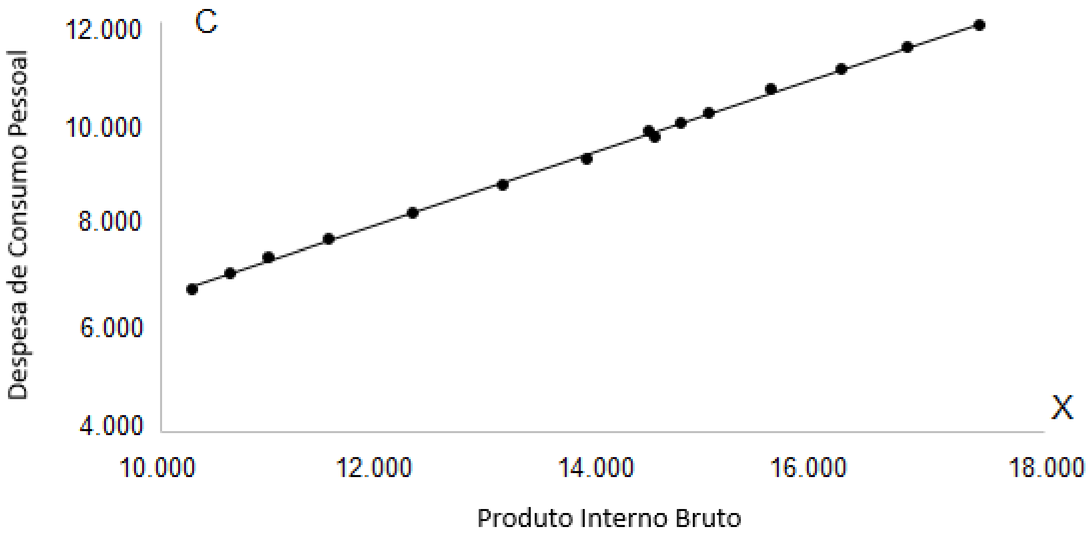
2.4.6 *Estimação dos parâmetros do modelo*

Depois que o econometrista obtém os dados, fornecidos pelo economista estatístico, será necessário distinguir e estimar os parâmetros do modelo econométrico.

Um método de estimação é consistente, se a função de estimação estatística (estimador) converge para o valor verdadeiro do parâmetro, quando o tamanho da amostra cresce [35].

Supondo que a relação linear entre as variáveis C e X (Tabela 4) é satisfatória, podemos estimar uma linha que melhor se ajusta aos pontos do gráfico de dispersão (a definição será dada no capítulo 3) e a resolução das inferências estatísticas. O problema de estimar os parâmetros β_0 e β_1 é o mesmo que ajustar esta reta, como podemos ver na figura 11. O Método dos Mínimos Quadrados é uma eficiente estratégia de estimação dos parâmetros da regressão e sua aplicação não é limitada apenas às relações lineares, como veremos no capítulo 3.

Figura 11: Despesas de Consumo Pessoal (C) e Produto Interno Bruto (X) dos EUA no período 2000-2014, em bilhões de dólares



Com base na tabela 4 elaboramos o gráfico (figura 11) através do Excel e obtemos os estimadores dos parâmetros de $\widehat{\beta_0} = -536,01$ e $\widehat{\beta_1} = 0,7169$. Portanto, a função consumo estimada é:

$$\widehat{C}_t = -536,01 + 0,7169X_t \tag{2.10}$$

O acento circunflexo em cima do $\widehat{C}$ (lê-se C-chapéu) indica, por convenção, que se trata de um valor estimado.

Como indica a Figura 11, a linha de regressão se ajusta muito bem, pois os pontos, que são valores reais ficam bem próximos da linha de regressão, isto significa que este modelo apresenta um bom ajuste, mas não quer disser que devemos aceita-lo sem testar. O coeficiente angular (a PMC) encontrado foi de 0,7169 indicando que no período analisado, um aumento de um dólar na renda disponível acarretará um aumento aproximadamente de 72 centavos nas despesas reais de consumo [14].

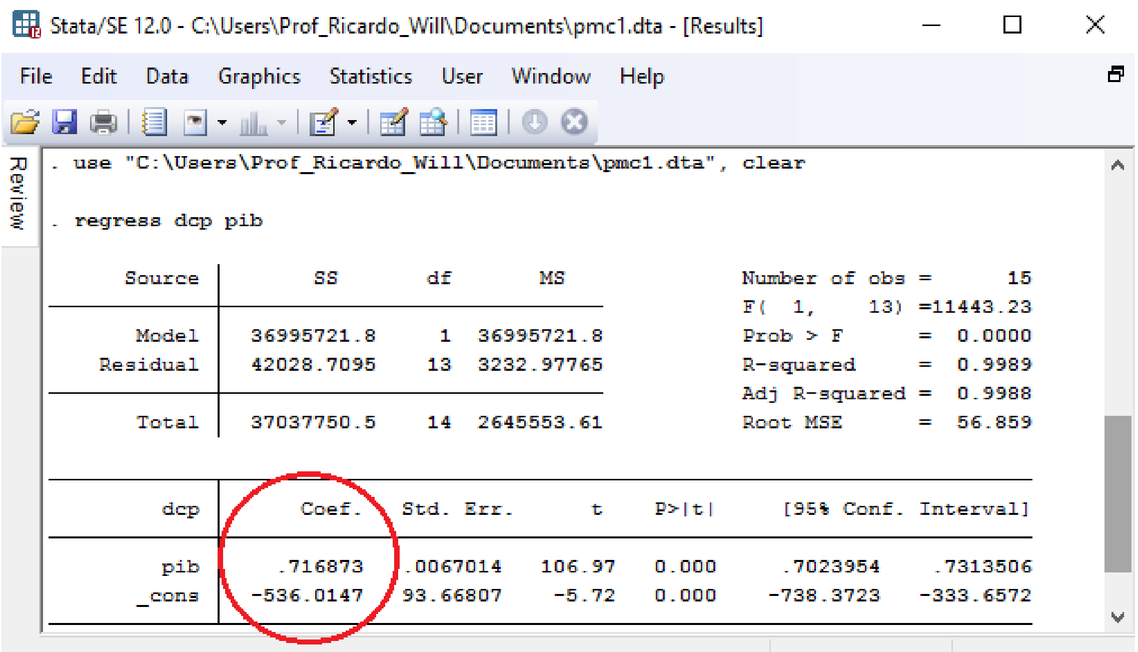
2.4.7 *Teste de hipóteses*

O econometrista depois de encontrar os parâmetros no modelo econométrico faz os testes de hipóteses graça ao modelo que ele obtém da estatística matemática e do limite de confiança e assim terá condições de validar ou refutar a teoria.

No nosso exemplo, a PMC foi de 0,7169 e de acordo com a teoria keynesiana esperávamos um valor 0<PMC<1.

Com auxílio do programa Stata 12.0 verificamos que os parâmetros no modelo econométrico são: $\widehat{\beta_0} = \mathbf{-536,0147}$ e $\widehat{\beta_1} = \mathbf{0,716873}$ (Figura 12).

Figura 12: Stata 12.0 - Parâmetros no modelo econométrico



2.4.8 *O modelo é adequado?*

No caso da validação da hipótese o econometrista terá condições de utilizar este modelo e substituir os valores das variáveis e fazer projeção ou previsão de valores futuros para a variável de interesse.

De acordo com o Economic Report of the President, 2015 a economia dos EUA cresceu 2,4% em 2015. A taxa é a mesma registrada no ano anterior. O PIB foi de 17.937,8 bilhões de dólares e a propensão marginal a consumir de 12.267,9 bilhões de dólares [7].

Substituindo o valor de X_{2016} = 17.937,8 bilhões de dólares na equação (2.10) temos:

$$\widehat{C}_{2016} = -536,01 + 0,7169X_{2016}$$
$$\widehat{C}_{2016} = -536,01 + 0,7169.(17937,8)$$
$$\widehat{C}_{2016} = -536,01 + 12859,61$$
$$\widehat{C}_{2016} = 12323,6$$

Comparando com o valor real de 12.267,9 com o previsto de 12.323,6, percebemos que as despesas médias previstas no modelo na equação (2.10) foi maior que o registrado em 2015 em 55,7 bilhões de dólares.

Este erro é "grande"ou é "pequeno"? Existem fatores que podem ter contribuído como por exemplo: aumento nos impostos ou taxas dos produtos, desemprego e outros fatores, que não foram considerados. Como já foi dito anteriormente a econometria é uma ciência social, portanto não existe uma precisão, sempre em uma previsão os erros são inevitáveis, mas o correto é que sejam o mínimo possível

Um exemplo de uma aplicação macroeconômica do modelo econométrico utilizando este modelo poderia ser:

O Governo Federal decide corrigir a tabela das alíquotas do imposto de renda, diminuir o imposto sobre as despesas de consumo e sobre o emprego. Qual seria o efeito na economia [14]?

É claro que teríamos um impacto na economia em relação ao consumo, ao nível de emprego e ao investimento, que teriam um aumento significativo. Pela teoria macroeconomia devemos utilizar o conceito de **multiplicador da renda** (M).

O multiplicador de renda é efeito um propagador de renda gerado por um investimento inicial, onde lembrando da equação 2.6 que: 1 - PMC = PMS, portanto o efeito multiplicador é o inverso da PMS.

$$M = \frac{1}{1 - PMC} \tag{2.11}$$

utilizando o valor de PMC = 0,7169, obtida através dos dados empíricos para a economia dos EUA para o período de 2000-2014, este multiplicador será de:

$$M = \frac{1}{1 - 0,7169} \approx 3,53$$

Isto significa que a cada um dólar gasto por uma pessoa nos EUA tem a capacidade de transformar em cerca de 3,53 dólares de renda disponível em circulação no país.

Vamos ilustrar o multiplicador de renda em uma situação utilizando a PMC de 0,7169. Um turista chega em uma cidade e paga a hospedagem de hotel com $100,00. O dono do hotel paga uma dívida com seu fornecedor de carnes de $71,69. Este fornecedor paga sua dívida com o criador de suínos de $51,40. O criador paga ao veterinário que cuida dos animais $36,84. O veterinário paga a recepcionista $26,41. A recepcionista paga a mercearia $18,93. A mercearia paga ao entregador $13,57. O entregador paga $9,72 a lanchonete e o dinheiro continua circulando pela cidade, pois os vários setores dentro de uma economia estão correlacionados, podemos verificar o efeito multiplicador total deste exemplo na Tabela 26 do Apêndice A.

Quanto menos a sociedade poupar (PMS), maior será o efeito multiplicador. Enquanto mais ela poupar, menor será o impacto do aumento do investimento (ΔI) no aumento da renda (PIB).

O exemplo mostra o efeito multiplicador como um efeito cascateador de dinheiro, portanto se ocorrer um investimento inicial de $100,00, isto acarretará na economia uma renda circulante (aumento da renda) de $353,00.

$$\Delta X = \Delta I.M = 100,00.3,53 = \$353,00$$

Este controle tanto poderá se dar no âmbito macroeconômico (política) ou no âmbito microeconômico (dentro de uma empresa ou por exemplo, o comportamento de consumo de cidadãos). E para controle de política apenas no nível macroeconômico, por exemplo renda, emprego, produto, crescimento da economia, e tantas outras variáveis que existe inflação.

2.5 EXERCÍCIOS SOBRE CONCEITOS DE ECONOMETRIA

Os exercícios propostos possuem as devidas resoluções, portanto, cabe ao professor decidir se os alunos deverão resolver todos ou parcialmente ou apenas discutir as resoluções.

1. A tabela 5 contém informações sobre o consumo (C) e renda familiar disponível (X), em reais, para uma amostra de $n = 17$. Com base nestes dados, resolva os itens abaixo.

Tabela 5: Exercícios sobre consumo pessoal e renda disponível

Observações	**Consumo (em R$)**	**Renda (em R$)**
1	784,00	832,20
2	796,80	815,00
3	796,70	834,30
4	795,70	823,50
5	803,30	844,00
6	811,60	869,40
7	814,50	887,40
8	822,60	890,90
9	823,90	874,10
10	834,30	874,40
11	831,30	874,30
12	836,20	875,00
13	848,80	936,90
14	865,00	933,70
15	899,30	937,80
16	884,30	961,70
17	899,80	954,70

a-) Construa um diagrama de dispersão para analisar o comportamento do consumo em função da renda.

b-) Ajuste uma linha que melhor se ajusta aos pontos do diagrama de dispersão.

c-) Determine os estimadores dos parâmetros da função consumo (com auxílio do Excel ou Stata).

d-) Determine a função keynesiana que descreve o comportamento do consumo em função da renda (com o Excel ou Stata).

e-) Estime o valor esperado para o consumo de uma família com renda familiar de 1.500 reais.

f-) Estime o valor esperado para a renda de uma família com consumo de 1.500 reais.

g-) Determine a propensão marginal ao consumo (PMC) e a propensão marginal a poupar (PMS).

h-) Determine o multiplicador de renda keynesiano (M).

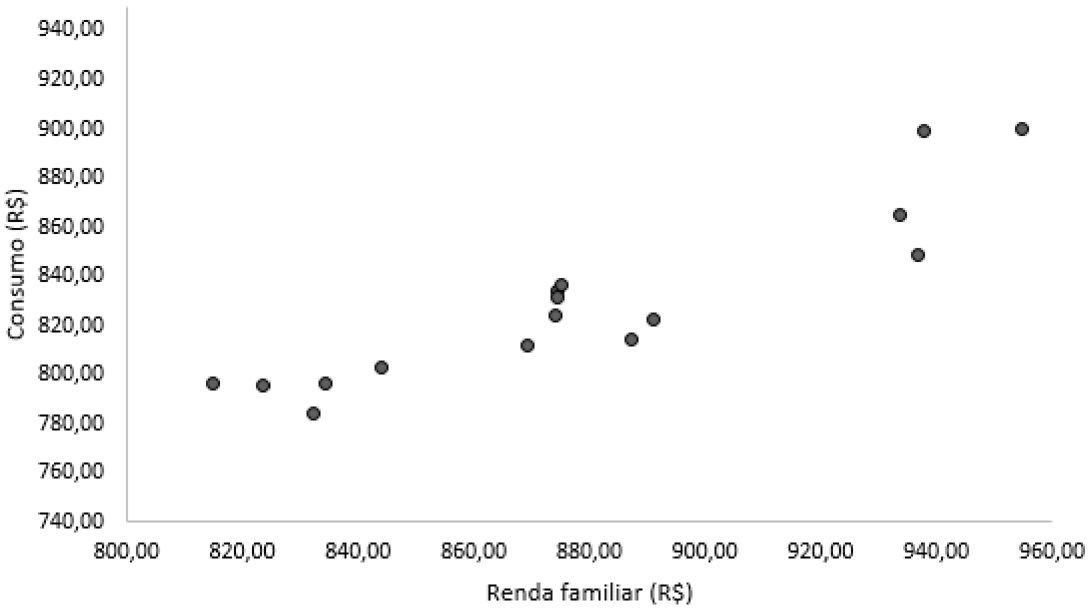
Resolução:

Neste exercício utilizaremos o Excel para que todos os alunos do Ensino Médio tenham condições de resolver.

a-) Construa um diagrama de dispersão para analisar o comportamento do consumo em função da renda.

Ver gráfico da Figura 13.

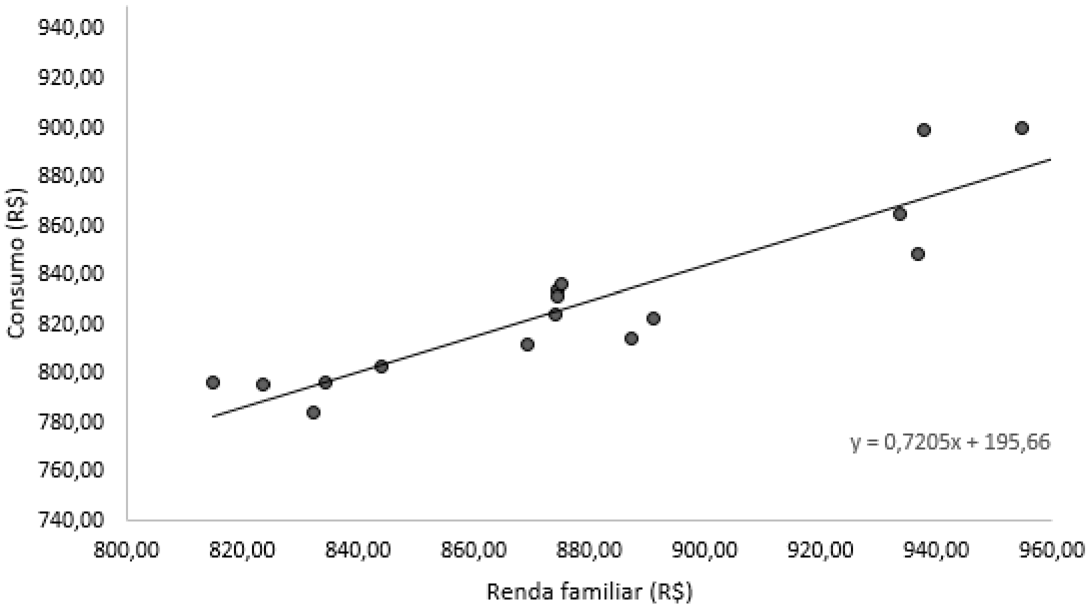
Figura 13: Gráfico de dispersão do consumo em função da renda



b-) Ajuste uma linha de regressão no diagrama de dispersão.

Ver gráfico da Figura 14.

Figura 14: Gráfico de dispersão com a reta de regressão do consumo em função da renda



c-) Determine os estimadores dos parâmetros da função consumo (com auxílio do Excel ou Stata).

Com base na reta de regressão do gráfico obtemos os parâmetros de α = 195,66 e β = 0,7205.

d-) Determine a função keynesiana que descreve o comportamento do consumo em função da renda (com o Excel ou Stata).

$$\hat{C} = 0,7205X + 195,66$$

e-) Estime o valor esperado para o consumo de uma família com renda familiar de 1.500 reais.

$$\begin{aligned} \hat{C} &= 0,7205 \cdot 1500 + 195,66 \\ \hat{C} &= 1080,75 + 195,66 \\ \hat{C} &= 1276,41 \end{aligned}$$

Portanto, o valor esperado para o consumo será de R$1.276,41 para uma renda de R$1.500,00.

f-) Estime o valor esperado para a renda de uma família com consumo de 1.500 reais.

$$\begin{aligned} 1500 &= 0,7205X + 195,66 \\ 1304,34 &= 0,7205X \\ \hat{C} &= 1810,33 \end{aligned}$$

Portanto, o valor esperado para a renda familiar será de R$1.810,33 para um consumo de R$1.500,00.

g-) Determine a propensão marginal ao consumo (PMC) e a propensão marginal a poupar (PMS).

A PMC é o nosso parâmetro β, portanto: PMC = 0,7205 = 72,05%
A PMS é dada pela equação:

$$\begin{aligned} PMS &= 1 - PMC \\ PMS &= 1 - 0,7205 \\ PMS &= 0,2975 = 29,75\% \end{aligned}$$

h-) Determine o multiplicador de renda keynesiano (M).

O multiplicador de renda (M) é o inverso da PMS, então:

$$M = \frac{1}{PMS} = \frac{1}{0,7205}$$
$$M \approx 3,58$$

2. (APO/MPOG – ESAF/2002) Com relação ao multiplicador keynesiano, é correto afirmar que:

A) se a propensão marginal a consumir for igual à propensão marginal a poupar, o seu valor será igual a um.

B) numa economia fechada, seu valor depende da propensão marginal a poupar, pode ser menor do que um e só é válido para os gastos do governo.

C) numa economia aberta seu valor depende da propensão marginal a consumir e importar, pode ser negativo e vale apenas para os gastos do governo e exportações autônomas.

D) numa economia fechada, seu valor depende da propensão marginal a poupar, não pode ser menor do que um e vale para qualquer componente dos denominados gastos autônomos agregados.

E) seu valor para uma economia fechada é necessariamente menor do que para uma economia aberta.

Resolução: D

3. O que é propensão marginal ao consumo? E propensão marginal a poupar? Qual a relevância destes conceitos dentro da macroeconomia keynesiana [17]?

   Resolução:
   Propensão Marginal ao Consumo (PMC) é um valor que varia entre 0 e 1, que expressa o comportamento dos agentes de uma determinada economia em relação a renda obtida. Quanto mais propensos estes forem a consumir, maior será o valor da PMC. Por exemplo, numa economia em que seus agentes estão dispostos a comprometer 70% da sua renda com o consumo, a PMC é igual a 0,7. Por tabela, a Propensão Marginal à Poupar (PMS) representa a parcela da renda dos agentes econômicos de uma determinada economia a qual estes comprometem-se a poupar (isto é, a não gastar). Portanto, resgatando o exemplo anterior, se a PMC é 0,7, a PMS será 0,3 – totalizando, PMC + PMS = 1 = 100% da renda disponível.

   O conceito de PMC é de grande valia quando observa-se o multiplicador keynesiano (M = 1/1-c), , em que "c" = PMC, do qual este é uma das variáveis. Quanto maior for o seu valor, maior será a renda final da economia em análise - o consumo (demanda) estimula a produção, que emprega cada vez mais os meios de produção disponíveis, empregando mais pessoas, que consumirão mais, formando um ciclo virtuoso.

   Nesta mesma fórmula, observa-se também que o multiplicador é determinado pelo inverso da PMS, sendo $1 - c = s \rightarrow M = 1/s$. Logo, quanto menor for o seu valor, considerando ser este um número compreendido entre 0 e 1, maior será o multiplicador, em que "s" = PMS.

4. Como calcular propensão marginal ao consumo (PMC), propensão marginal a poupar (PMS) e o multiplicador de renda keynesiano (M) [8]?

Resolução:

I Defina as variáveis. Essa é uma conta fácil de fazer. A parte difícil e acertar os valores. Vamos assumir que uma pessoa tenha um salário anual de 50 mil reais. Todo ano, o seu salário sobe 10 mil reais e seus gastos sobem pela 80% do que você passou a ganhar. Essa fração é a PMC.

II Essa fração adicional de quanto às pessoas podem gastar é a propensão marginal a consumir. Ela pode ser computada através da divisão da variação do consumo pela variação de renda disponível:

$$PMC = \frac{\Delta C}{\Delta X}$$

III Conforme o passo acima, o seu salário para os anos 1, 2, 3, 4 e 5 é de 50, 60, 70, 80 e 90 mil, respectivamente. Assim, o consumo para os ano 1, 2, 3, 4, 5 é de 42, 50, 58, 66, 74 mil respectivamente. Portanto, se sua renda sobe em 10 mil reais, seus gastos sobem em 8 mil. A equação da PMC é:

$$PMC = \frac{8.000}{10.000} = 0,8$$

IV Economistas também gostam avaliar a propensão marginal a poupar, que é a fração de renda que as pessoas guardam. Já que uma pessoa pode optar entre gastar ou poupar sua renda adicional, a soma entre a propensão marginal a poupar e a propensão marginal a consumir deve ser sempre 1.

V Calculando a PMS vamos utilizar $PMS = 1 - PMC$. Então teremos:

$$PMS = 1 - 0,8 = 0,2$$

VI Calculando o multiplicador de renda keynesiano que é o inverso da PMS, teremos:

$$M = \frac{1}{PMS} = \frac{1}{0,2} = 5$$

5. Dada a função de consumo:

$$C = \frac{5\left(2\sqrt{X^3}+3\right)}{X+10}$$

Determine a propensão marginal ao consumo (PMC), a propensão marginal a poupar (PMS) e o multiplicador de renda (M), quando X = 100.

Resolução: A PMC é definida como a derivada do consumo em relação a renda conforme a equação (2.2) e que $0 < PMC < 1$.

$$PMC = \frac{dC}{dX}$$

Utilizando a regra do quociente teremos:

$$PMC = \frac{(X+10)\frac{d}{dX}\left[5\left(2X^{\frac{3}{2}}+3\right)\right] - 5\left(2X^{\frac{3}{2}}+3\right)\frac{d}{dX}(X+10)}{(X+10)^2}$$

$$PMC = \frac{(X+10)\left[5\left(3X^{\frac{1}{2}}\right)\right] - 5\left(2X^{\frac{3}{2}}+3\right)(1)}{(X+10)^2}$$

$$PMC = \frac{(X+10)\left(15\sqrt{X}\right) - \left(10\sqrt{X^3}+15\right)}{(X+10)^2}$$

Substituindo o valor de X temos:

$$PMC = \frac{(100+10)\left(15\sqrt{100}\right) - \left(10\sqrt{100^3}+15\right)}{(100+10)^2}$$

$$PMC = \frac{(110 \cdot 150) - (10 \cdot 1000 + 15)}{(110)^2}$$

$$PMC = \frac{16500 - 10015}{12100} = \frac{6485}{12100}$$

$$PMC = \frac{6485}{12100} \approx 0,536 = 53,6\%$$

A propensão a poupar (PMS) é:

$$PMS = 1 - PMC$$

conforme equação (2.6)

$$PMS = 1 - 0,536 = 0,464 = 46,4\%$$

O multiplicador de renda (M) é o inverso da PMS, teremos:

$$M = \frac{1}{PMS} = \frac{1}{0,464} \approx 2,16$$

Portanto: PMC = 0,536 = 53,6%, a PMS = 0,464 = 46,4% e M = 2,16

3

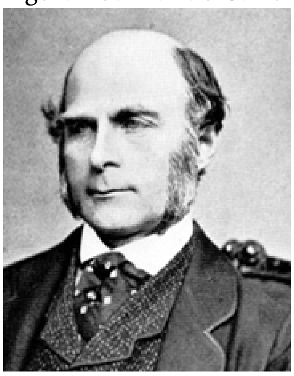
MODELO DE REGRESSÃO LINEAR SIMPLES

Neste capítulo vamos tratar do modelo de regressão linear simples (MRLS) que descreve a relação entre duas variáveis, que pode ser representada por uma linha reta.

3.1 A LEI DA REGRESSÃO UNIVERSAL DE GALTON

Francis Galton [1]. procurou construir uma teoria puramente estatística da relação entre pais e filhos com modelos teóricos de herança, como por exemplo a estatura [3].

Figura 15: Francis Galton



1 Francis Galton (1822-1911), nascido na Inglaterra, foi um antropólogo, meteorologista, matemático e estatístico. Destacou-se por criar o conceito de correlação e a aplicar métodos estatísticos para o estudo das diferenças e herança humanas de inteligência, e introduziu a utilização de questionários e pesquisas para coletar dados sobre as comunidades humanas, o que ele precisava para obras genealógicas e biográficas e para os seus estudos antropométricos. Era primo de Charles Darwin e, baseado em sua obra, criou o conceito de "eugenia", teoria que busca produzir uma seleção nas coletividades humanas, baseada em leis genéticas. O primeiro livro importante para o pensamento de Galton foi Hereditary Genius (1869) [27]

A característica que uma criança possuía não teria vindo necessariamente de um dos pais, mas poderia ter sido transmitida de forma latente pelas gerações e se manifestado naquele indivíduo [12].

Galton sabia que o ponto para o qual a regressão tendia não poderia ser fixado, mas o que poderia ser testado, pois as variações se davam através de mudanças lentas e graduais ou abruptas, que coincidiriam com mudanças no equilíbrio orgânico e que pudessem ser transmitidas hereditariamente. Ele também considerava a impossibilidade de tirar a média entre homem e mulher a fim de avaliar a contribuição hereditária de cada um [12].

Ele afirmou que a sua lei da hereditariedade envolvia cinco constantes que podiam ser determinadas separadamente, mas que eram conectadas por uma equação. Esta equação dependeria dos fatos ocorridos em gerações sucessivas da mesma população [12]. Galton observou que quanto maior a altura de um grupo de pessoas comparado a outro grupo de menor estatura, a altura de seus descendentes tende a ser maior, no entanto menores que seus pais (em média). Dessa forma, os descendentes de pais altos, naturalmente também são altos, mas não tanto quanto seus pais. Os descendentes de pais baixos, naturalmente também são baixos, mas não tanto quanto seus pais. O aumento da variabilidade da população como um todo seria contrabalanceado pela regressão, o que faria com que ela entrasse em equilíbrio, isto para ele era uma "regressão à mediocridade" [14].

3.2 A INTERPRETAÇÃO MODERNA DA REGRESSÃO

Regressão é um dos mais versáteis e populares procedimentos estatísticos.

A análise de regressão é frequentemente utilizado no mundo dos negócios e do investimento para tentar prever o efeito de certas entradas em uma saída.

A análise de regressão é o método mais importante da econometria [16].

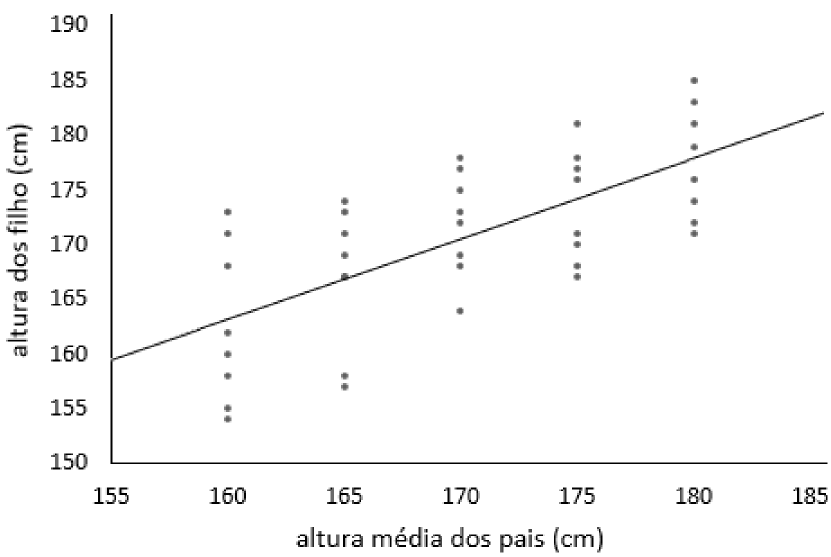
Por exemplo, um analista pode querer tentar prever o efeito do preço do aço sobre as vendas de automóveis, ou uma empresa pode querer ver se as suas vendas podem ser previstos pelo movimento no PIB.

Na Tabela 6, temos uma distribuição das alturas dos filhos em relação a altura média dos pais, em uma população hipotética, podemos observar pela linha de regressão do gráfico (Figura 16), que a altura média dos filhos aumenta com o aumento da altura média dos pais, conforme a teoria de Galton. [14]

Tabela 6: Altura dos filhos em relação a altura média dos pais (em cm)

alturas (em cm)								
média dos pais	**filhos**	**filhos**	**filhos**	**filhos**	**filhos**	**filhos**	**filhos**	**filhos**
160	155	154	158	160	162	168	171	173
165	157	158	167	169	173	167	171	174
170	164	168	169	172	173	175	177	178
175	167	168	170	171	176	177	178	181
180	171	172	174	176	179	181	183	185

Figura 16: Altura dos filhos em relação a altura média dos pais



Ao escrever um modelo que explicará a relação entre duas variáveis devemos conhecer os efeitos que algumas variáveis exercem, ou que parecem exercer, sobre outras [16].

O principal objetivo da regressão é predizer ou explicar o comportamento de uma variável dependente (**Y**) usando uma ou mais variáveis conhecidas, como as variáveis independentes (**X**). Se somente uma variável é usada na regressão, procede-se à chamada **regressão linear simples** e caso seja utilizada duas ou mais variáveis independentes no modelo, representado por uma equação, se chamada de **regressão linear múltipla** [36].

3.3 TERMINOLOGIA E NOTAÇÃO

As variáveis X e Y têm vários nomes diferentes, os quais são intercambiáveis, no entanto os termos: variável dependente e variável independente são usados com mais frequência em econometria. Essa terminologia está resumida na tabela (7) [38].

Tabela 7: Terminologia para a regressão simples

Y	**X**
variável dependente	variável independente
variável explicada	variável explicativa
variável de resposta	variável de controle
variável prevista	variável previsora
regressando	regressor

Como já vimos no capítulo anterior, na equação 2.9, temos:

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{3.1}$$

A Tabela 8 apresenta, resumidamente, o significado prático de cada componente da equação de regressão 3.1 [36].

Tabela 8: Significado dos termos de uma equação de regressão simples

Notação	**Terminologia técnica**	**Significado prático**
Y	Variável dependente	Representa o comportamento de Y quando são assumidos certos valores por parte de X.
β_0	Intercepto	Representa o valor de Y quando X = 0.
β_1	Efeito marginal de X sobre Y	Expressa o impacto da variação de uma unidade da variável X sobre a variável Y.
ε	Termo de erro	Variável que inclui todos os fatores residuais e os possíveis erros de medição.

Uma questão difícil é saber se o modelo (Equação 3.1) nos permite tirar conclusões ***ceteris paribus***[2]. sobre como X afeta Y.

2 Ceteris paribus é usada na economia e finanças como uma indicação abreviada do efeito de uma economia variável em outra, mantendo todas as outras variáveis constantes, sem que as demais variáveis sofram alterações. Por exemplo, considere a Lei da Oferta e da Procura. A demanda (procura) por um determinado produto é superado pelo fornecimento do mesmo, os preços provavelmente irão subir [20]

Um modelo seria perfeito se todos os pontos pertencessem a reta de regressão, mas na realidade isto não ocorre, portanto podemos minimizar o erro através do método conhecido como Mínimos Quadrados Ordinários

Regressão também é utilizada para determinar covariância e correlação, que são variáveis usadas no mundo dos investimentos para mostrar o quão perto duas variáveis tendem a mover-se na mesma direção ou em direções diferentes. Esta é uma informação importante para os investidores que desejam diversificar as ações que não estão correlacionadas com os que já possuem [21].

3.4 DIAGRAMA DE DISPERSÃO

Análise de correlação tem como principal objetivo medir a força ou o grau de associação linear entre duas variáveis [14].

Na prática temos a necessidade de estudar a relação entre duas variáveis, que é feito através do Diagrama de Dispersão.

Gráfico ou Diagrama de Dispersão é o método gráfico feito sobre dois eixos, x e y, que mostra a relação entre duas variáveis quantitativas. Elas estão relacionadas se a mudança de uma provoca a mudança na outra.

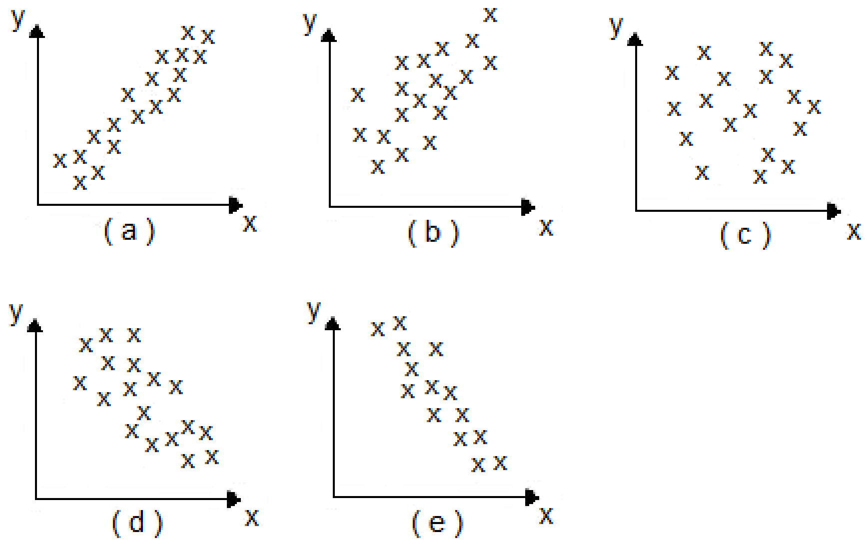
Exemplos:

- Velocidade x Consumo de combustível.
- Número de peças produzidas x Número de peças defeituosas.
- Salário x Tempo de escolaridade.
- Despesas de consumo x Renda disponível.

Observando o diagrama de dispersão podemos verificar a concentração dos pontos e sua tendência determinando se existe correlação entre as variáveis (Figura 17). Inicialmente observamos:

- Direção.
- Forma (linear, não linear, aglomerados).
- Pontos discrepantes.

Figura 17: Possíveis padrões para Diagramas de Dispersão.



Legendas da Figura 17:

(a) Elevada correlação positiva.

(b) Moderada correlação positiva.

(c) Ausência de correlação.

(d) Moderada correlação negativa.

(e) Elevada correlação negativa.

A análise visual do gráfico da relação entre variáveis nem sempre é confiável, pois a intensidade da concentração de uma relação linear depende muito da escala utilizada nos eixos, portanto devemos utilizar uma medida numérica para suplementar o gráfico, isto é, o coeficiente de correlação linear (r) ou coeficiente de correlação de Pearson, que veremos mais adiante, neste capítulo.

3.5 ESTIMAÇÃO DOS PARÂMETROS PELO MÉTODO DOS MÍNIMOS QUADRADOS ORDINÁRIOS

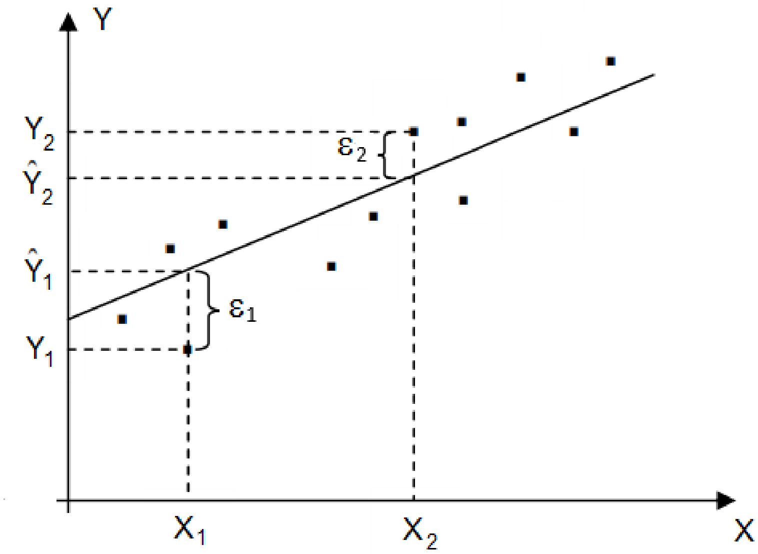
Nosso objetivo é ajustar uma linha que coincida com as observações tanto quanto possível no gráfico de dispersão. Para tornar operacional esta abordagem temos que determinar a medida de grau de coincidência, portanto podemos escrever as observações de Y em termos da regressão Y sobre X. [1] Porém, qualquer linha traçada através da dispersão poderá ser representada como

$$\widehat{Y}_i = \widehat{\beta_0} + \widehat{\beta_1} X_i + \varepsilon_i \tag{3.2}$$

onde, $\widehat{Y}_i$ é o valor de Y dado por essa linha para cada valor de X_i. A diferença entre Y_i (o valor de Y) e $\widehat{Y}_i$, pode ser denotada por ε_i (Figura 18), de modo que [33]

$$|Y_i - \widehat{Y}_i| = |Y_i - \widehat{\beta_0} - \widehat{\beta_1} X_i| = |\varepsilon_i|$$

Figura 18: O termo de erro



Nosso objetivo é determinar os valores de $\widehat{\beta_0}$ e $\widehat{\beta_1}$ que minimizem a soma dos quadrados dos erros.

$$\sum |\varepsilon_i| = 0$$

3.5.1 *Cálculo do parâmetro* $\widehat{\beta_0}$

Linhas diversas traçadas através da dispersão terão diferentes $\widehat{\beta_0}$ e $\widehat{\beta_1}$, apesar de os X_i e Y_i, permanecerem inalterados, e assim resultarão em valores diversos de ε_i. Através do método Método dos Mínimos Quadrados Ordinários (MQO) temos a melhor linha que atravessa a dispersão que minimiza a **soma dos quadrados dos resíduos (ou erros) $\boldsymbol{\varepsilon_i}$**. Denotando esta soma por **SQR** teremos [1]:

$$SQR = \sum_{i=1}^{n} \varepsilon_i^2 \tag{3.3}$$

$$SQR = \sum_{i=1}^{n} (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i)^2 \tag{3.4}$$

Uma vantagem dessa última definição é que devido ao fato de não necessitarmos nos preocupar com o sinal de ε. Para obter o mínimo valor de SQR diferenciamos SQR tanto em relação a $\widehat{\beta}_0$, quanto a $\widehat{\beta}_1$. Primeiramente vamos derivar SQR em relação a $\widehat{\beta}_0$.

$$\frac{d(SQR)}{d\widehat{\beta}_0} = \sum_{i=1}^{n} \frac{d}{d\widehat{\beta}_0} (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i)^2$$

$$\frac{d(SQR)}{d\widehat{\beta}_0} = \sum_{i=1}^{n} 2\left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i\right) \cdot (-1)$$

$$\frac{d(SQR)}{d\widehat{\beta}_0} = -2\sum_{i=1}^{n}\left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i\right) \tag{3.5}$$

Igualando a equação (3.5) a zero, temos

$$\sum_{i=1}^{n}\left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i\right) = 0 \tag{3.6}$$

Simplificando-se, estes se reduzem às chamadas *equações normais* para a linha da reta, onde n representa o número de observações.

$$\sum_{i=1}^{n} Y_i - n\widehat{\beta}_0 - \widehat{\beta}_1 \sum_{i=1}^{n} = 0$$

$$\sum_{i=1}^{n} Y_i = n\widehat{\beta}_0 + \widehat{\beta}_1 \sum_{i=1}^{n} X_i \tag{3.7}$$

Dividindo a equação (3.7) por n, onde $\bar{Y}$ e $\bar{X}$ são as médias aritméticas de X_i e Y_i, respectivamente obtemos

$$\frac{\sum_{i=1}^{n} Y_i}{n} = \frac{n\widehat{\beta}_0}{n} + \frac{\widehat{\beta}_1 \sum_{i=1}^{n} X_i}{n}$$

$$\bar{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 \bar{X} \tag{3.8}$$

Isolando $\widehat{\beta_0}$ podemos determinar o valor deste estimador.

$$\therefore \quad \widehat{\beta_0} = \bar{Y} - \widehat{\beta_1}\bar{X} \tag{3.9}$$

3.5.2 *Cálculo do parâmetro $\widehat{\beta_1}$*

Com a intenção de simplificar a expressão de SQR antes de calcularmos sua derivada com relação a $\hat{\beta_1}$, sabemos que a linha dos mínimos quadrados deve passar através do ponto $(\bar{X}_i, \bar{Y}_i)$. Substituindo a equação (3.9) em (3.2) teremos

$$\widehat{Y_i} = (\bar{Y} - \widehat{\beta_1}\bar{X}) + \widehat{\beta_1}X_i + \varepsilon_i$$
$$\widehat{Y_i} - \bar{Y} = \widehat{\beta_1}(X_i - \bar{X}) + \varepsilon_i$$

Usamos letras minúsculas (x e y) para denotar desvios com relação à média, teremos

$$x_i = \widehat{X_i} - \bar{X} \qquad e \qquad y_i = \widehat{Y_i} - \bar{Y} \tag{3.10}$$

Portanto a equação (3.2) pode ser reescrita como

$$y_i = \widehat{\beta_1}x_i + \varepsilon_i$$
$$\varepsilon_i = y_i - \widehat{\beta_1}x_i \tag{3.11}$$

Substituindo a equação (3.11) em (3.3) e derivando SQR em relação a $\widehat{\beta_1}$

$$SQR = \sum_{i=1}^{n}\varepsilon_i^2$$
$$SQR = \sum_{i=1}^{n}(y_i - \widehat{\beta_1}x_i)^2$$
$$\frac{d(SQR)}{d\hat{\beta_1}} = -2\sum_{i=1}^{n}[x_i(y_i - \widehat{\beta_1}x_i)]$$

Igualando-se essa equação a zero e isolando $\hat{\beta_1}$ obtemos

$$\sum_{i=1}^{n}x_i(y_i - \widehat{\beta_1}x_i) = 0$$
$$\sum_{i=1}^{n}x_iy_i = \widehat{\beta_1}\sum_{i=1}^{n}x_i^2$$
$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}^{n}x_i^2} \tag{3.12}$$

Lembrando que x e y são os desvios com relação a média equações (3.10) podemos reescrever a equação (3.12).

$$\therefore \quad \widehat{\beta_1} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{3.13}$$

Portanto, podemos calcular os valores dos parâmetros $\widehat{\beta_1}$ e $\widehat{\beta_0}$ de uma equação de regressão linear simples utilizando as equações (3.13) e (3.9), receptivamente.

Também podemos utilizar a **notação matricial** para determinar $\widehat{\beta_0}$ e $\widehat{\beta_1}$, como veremos no exemplo 2 do capítulo 4.

Ao usar um modelo de regressão para fazer previsões devemos fazer somente dentro do intervalo de confiança.

3.6 NOTAÇÕES ESPECIAIS

i) $S_{XX} = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$

ii) $S_{XY} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}$

iii) $S_{YY} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2$

iv) $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \dfrac{S_{XY}}{S_{XX}}$

v) $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = \bar{Y} - \dfrac{S_{XY}}{S_{XX}}\bar{X}$

3.7 ESTIMAÇÃO DOS RESÍDUOS

A análise dos resíduos, ao se examinar o quanto e por que as observações se desviam da linha de regressão, é geralmente tão rica em informações quanto a própria linha de regressão. Entretanto, o fato de a linha de regressão ser obtida minimizando-se a soma dos quadrados dos resíduos (Equação 3.3).

Os resíduos $\varepsilon_i = Y_i - \widehat{Y}_i$ (ver Figura 18) são empregados na estimação de σ^2, portanto σ^2 são simplesmente as diferenças entre os valores Y reais e estimados.

O valor médio do resíduo ε_i é zero e não são correlacionados com X_i e $\widehat{X}_i$ (independência) [30]. Utilizando a equação (3.6) teremos:

$$\sum_{i=1}^{n} \varepsilon_i = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i) = 0$$

A soma dos valores observados Y_i é igual a soma dos valores ajustados $\widehat{Y}_i$.

$$\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \widehat{Y}_i = 0$$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \widehat{Y}_i$$

3.7.1 *Estimador Quadrado Médio dos Resíduos e o Desvio Padrão*

A variância de σ^2 dos termos de erro ε_i precisa ser calculada, assim como dos parâmetros β_0 e β_1, portanto será necessário em virtude das inferência a respeito da função de regressão e da predição de Y requerem uma estimativa de σ^2 [28].

$$SQR = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - (\widehat{\beta_0} + \widehat{\beta_1} x_i))^2$$

SQR é um estimador de σ^2, isto é,

$$E(SQR) = \sigma^2(n-2) = S_{YY} - \widehat{\beta_1} \cdot S_{XY} \tag{3.14}$$

Como o estimador σ^2 é desconhecido, devemos obter uma estimativa a partir dos dados, e então substituir por $\widehat{\sigma}^2$, que é um estimador não viciado, denominado de Estimador Quadrado Médio dos Resíduos (QMR) e é dado por

$$\widehat{\sigma}^2 = QMR = \frac{SQR}{n-2}$$

$$\widehat{\sigma}^2 = QMR = \frac{S_{YY} - \widehat{\beta}_1 \cdot S_{XY}}{n-2} \tag{3.15}$$

Onde a expressão $n - 2$ e conhecida como **número de graus de liberdade (gl)**, que representa o número de observações da amostra menos o número de restrições independentes impostas a ele.

O desvio padrão dos resíduos pode ser calculado através

$$\widehat{\sigma} = \sqrt{QMR} = \sqrt{\widehat{\sigma}^2} \tag{3.16}$$

Na prática substituímos σ^2 (desconhecido), pelo estimador consciente $\widehat{\sigma}^2$.

3.7.2 *Variância e Desvio Padrão dos Estimadores de MQO*

A variância é definida como o desvio quadrático médio da média (σ^2) e é calculada de uma amostra de dados como

$$var(X) = \sigma_X^2 = \frac{S_{XX}}{n} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \tag{3.17}$$

$$var(Y) = \sigma_Y^2 = \frac{S_{YY}}{n} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n} \tag{3.18}$$

O desvio padrão da variável (σ), por definição é igual a

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\frac{S_{XX}}{n}} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} \tag{3.19}$$

$$\sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{\frac{S_{YY}}{n}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}} \tag{3.20}$$

A covariância de XY (Cov_{XY}) significa co-variação, como as duas variáveis variam de forma conjunta e é definida

$$Cov_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n} = \frac{S_{XY}}{n} \tag{3.21}$$

Analisando a função de estimativas de mínimos quadrados dos dados amostrais precismos de alguma medida de confiabilidade ou precisão dos estimadores $\widehat{\beta_0}$ e $\widehat{\beta_1}$, pois os dados mudam de uma amostra para outra [14].

$$E(\widehat{\beta_1}) = \beta_1$$

$$E(\widehat{\beta_0}) = \beta_0$$

A variância dos estimadores β_0 e β_1 é igual a

$$var(\widehat{\beta_0}) = \sigma^2_{\widehat{\beta_0}} = \frac{\sum X_i^2}{n \sum x_i^2}\sigma^2 = \frac{\sum X_i^2}{nS_{XX}}\sigma^2 = \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}\right) \tag{3.22}$$

$$var(\widehat{\beta_1}) = \sigma^2_{\widehat{\beta_1}} = \frac{\sigma^2}{\sum x_i^2} = \frac{\sigma^2}{S_{XX}} \tag{3.23}$$

O desvio padrão dos estimadores $\widehat{\beta_0}$ e $\widehat{\beta_1}$, é

$$\sigma_{\widehat{\beta_0}} = \sqrt{\sigma^2_{\widehat{\beta_0}}} \tag{3.24}$$

$$\sigma_{\widehat{\beta_1}} = \sqrt{\sigma^2_{\widehat{\beta_1}}} \tag{3.25}$$

3.8 ANÁLISE DE VARIÂNCIA (ANOVA)

A Análise de Variância (ANOVA) é um procedimento utilizado para testar hipóteses e a importância de um ou mais fatores comparando as médias das variáveis de resposta em diferentes níveis dos fatores [32].

As análises ANOVA exigem dados com a suposição de que os erros tem distribuição normal.

A análise de variância, baseia-se na decomposição da soma de quadrados e nos graus de liberdade associados a variável resposta Y, portanto podemos dizer que o desvio de uma observação em relação à média pode ser decomposto como o desvio da observação em relação ao valor ajustado pela regressão mais o desvio do valor ajustado em relação à média.

$$(Y_i - \bar{Y}) = (Y_i - \bar{Y} + \widehat{Y}_i - \widehat{Y}_i) = (\widehat{Y}_i - \bar{Y}) + (Y_i - \widehat{Y}_i) \tag{3.26}$$

3.8.1 *Soma de Quadrados*

Utilizando a equação 3.26 e elevando cada componente ao quadrado e fazendo somatória de todo o conjunto de observações, teremos

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\widehat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 \tag{3.27}$$

Podemos verificar que a Soma de Quadrados Total (SQT) é decomposta pela Soma de Quadrados Explicada ou da Regressão (SQE) e Soma de Quadrados dos Resíduos (SQR) [38].

$$SQT = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \tag{3.28}$$

$$SQE = \sum_{i=1}^{n}(\widehat{Y}_i - \bar{Y})^2 \tag{3.29}$$

$$SQR = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 \tag{3.30}$$

Então podemos reescrever a equação 3.27 da seguinte forma:

$$SQT = SQE + SQR \tag{3.31}$$

$$SQT = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\widehat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 \tag{3.32}$$

3.8.2 *Graus de Liberdade*

Existe uma decomposição dos graus de liberdade associados (gl), onde $k = 1$ para o MRLS (k é quantidade de variáveis independentes). A decomposição é a seguinte:

$$\begin{cases} \text{para SQE, gl} = 1, \\ \text{para SQR, gl} = n \text{ - } (k + 1) = n \text{ - } (1 + 1) = n \text{ - } 2, \\ \text{para SQT, gl} = n \text{ - } 1. \end{cases}$$

3.8.3 *Quadrado Médio*

O quadrado médio é a divisão da soma de quadrados pelos respectivos graus de liberdade. Como os graus de liberdade são diferentes para cada, então a relação da decomposição da variabilidade não funcionará mais.

O Quadrado Médio Explicada ou da Regressão (QME) é

$$QME = \frac{SQE}{gl} = \frac{SQE}{1} = SQE = \sum_{i=1}^{n}(\widehat{Y}_i - \bar{Y})^2 \tag{3.33}$$

O Quadrado Médio dos Resíduos (QMR) é

$$QMR = \frac{SQR}{gl} = \frac{SQR}{n-2} = \frac{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{n-2} \tag{3.34}$$

conforme já visto na equação 3.15.

3.8.4 *Tabela de Análise de Variância*

Podemos resumir a ANOVA nesta tabela que ajudará a analisar os dados.

Tabela 9: Tabela Resumo ANOVA para o MRLS

Fonte	**gl**	**Soma de Quadrados**	**Quadrado Médio**
Regressão	1	$SQE = \widehat{\beta}_1 \sum_{i=1}^{n}(X_i - \bar{X})Y_i$	$QME = SQE$
Resíduo	n - 2	$SQR = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 - \widehat{\beta}_1 \sum_{i=1}^{n}(X_i - \bar{X})Y_i$	$QMR = \dfrac{SQR}{n-2}$
Total	n - 1	$SQT = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$	-

3.8.5 *Teste F*

Considerando o modelo de regressão linear simples, a análise de variância estabelece um teste estatístico para testar e avaliar o parâmetro β_1, isto é, testar as hipóteses [29].

Em um modelo, existe o interesse em testar as hipóteses.

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

O teste F (ou razão F) testa se duas variâncias são iguais. Em caso de variâncias idênticas, F = 1. Tal distribuição é tabelada.

Valores altos desta estatística fornecem evidências a favor da hipótese H_1, ou seja, as médias diferem.

Então, consideremos o **Teorema de Cochran**.

Sejam $Z_1,\ Z_2,\ ...,\ Z_k$ variáveis independentes com distribuição $N(0, 1)$.

$$\sum_{i=1}^{k} Z_i^2 \ \text{ possui distribuição } \ \chi^2_{(k)}.$$

Se tivermos

$$Q = Q_1 + Q_2 + ... + Q_k,$$

em que $Q_i\ ,\ i = 1, 2, ..., q\ \ (q \leq p)$ são somas de quadrados, cada um com p_i graus de liberdade, tal que

$$p = \sum_{i=1}^{q} p_i,$$

então obtemos que $Q_i \sim \chi^2_{(p_i)}$ e são independentes para qualquer $i = 1, 2, ..., q$. Sob H_0, $Y_i \sim N(\beta_0, \sigma^2)$. Então, segue pelo Teorema de Cochran que

$$\chi_T = \frac{SQT}{\sigma^2} \sim \chi^2_{(n-1)};$$

$$\chi_E = \frac{SQR}{\sigma^2} \sim \chi^2_{(n-2)} \text{ e}$$

$$\chi_R = \frac{SQE}{\sigma^2} \sim \chi^2_{(1)},$$

com as distribuições Qui-Quadradas χ_E e χ_R independentes.

Desta forma, propomos a estatística do teste

Como F_0 é uma proporção de duas variáveis χ^2, cada uma dividida pelos seus graus de liberdade, segue que $F_0 \sim F_{(1,n-2)}$.

Uma motivação, baseada nas esperanças dos quadrados médios sugere que valores grandes de F_0 levem a H_1 e valores de F_0 próximos de 1 levem a H_0. Logo, rejeitamos H_0 com um nível de significância α se $F_0 > F_{(1-\alpha,1,n-2)}$.

$$F_0 = \frac{\frac{\chi_E}{1}}{\frac{\chi_R}{n-2}} = \frac{\frac{SQE}{\sigma^2}}{\frac{SQR}{(n-2)\sigma^2}} = \frac{QME}{QMR}$$

$$F_0 = \frac{QME}{QMR} \tag{3.35}$$

3.9 COEFICIENTE DE CORRELAÇÃO LINEAR E DE DETERMINAÇÃO

3.9.1 *Coeficiente de Correlação Linear de Pearson*

O coeficiente de correlação de Pearson é uma medida da intensidade de associação linear entre duas variáveis, que é frequentemente usada para uma amostra de observações de duas variáveis [1].

A correlação é calculada independente da unidade de medida das variáveis independentes e dependente.

A técnica usada para calcular este coeficiente, supõe que a associação entre as variáveis seja linear.

O coeficiente de correlação, normalmente representado por $\boldsymbol{r}$, pode variar entre -1 e +1. Quando $r = 0$ significa que as duas variáveis não dependem linearmente uma da outra, entretanto pode existir uma dependência não linear, portanto deve ser investigado por outros meios.

Com base na tabela 10 podemos classificar o grau de correlação pelo valor obtido do coeficiente [24].

Tabela 10: Classificação do coeficiente de correlação

Valor do coeficiente	**Classificação**
$0,90 < \|r\| \leq 1,00$	Uma correlação muito forte
$0,70 < \|r\| \leq 0,89$	Uma correlação forte
$0,40 < \|r\| \leq 0,69$	Uma correlação moderada
$0,20 < \|r\| \leq 0,39$	Uma correlação fraca
$0,01 < \|r\| \leq 0,19$	Uma correlação bem fraca
$\|r\| = 0$	Uma correlação inexistente

Valores negativos do coeficiente de correlação indicam uma correlação do tipo inversa, isto é, quando x aumenta y diminui.

Valores positivos do coeficiente de correlação ocorrem quando x e y variam no mesmo sentido, isto é, quando x aumenta y aumenta ou quando x diminui y também diminui.

O coeficiente de correlação de Pearson pode ser calculado através

$$r = \frac{\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n}}{\sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}} \cdot \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n}}} \tag{3.36}$$

$$r = \frac{Cov_{XY}}{\sigma_X \cdot \sigma_Y} \tag{3.37}$$

$$r = \frac{S_{XY}}{\sqrt{S_{XX} \cdot S_{YY}}} \tag{3.38}$$

3.9.2 *Coeficiente de Determinação*

O coeficiente de determinação r^2 é amplamente utilizado para medir a qualidade do ajustamento de uma linha de regressão. Ele mede em que percentual a variável dependente pode ser explicada pela variável independente (e vice-versa) [11].

Este coeficiente indica basicamente o quanto o modelo foi capaz de explicar os dados coletados.

É definido elevando o coeficiente de correlação (3.38) de Pearson ao quadrado.

$$(r)^2 = \left(\frac{S_{XY}}{\sqrt{S_{XX} \cdot S_{YY}}} \right)^2$$

$$r^2 = \frac{(S_{XY})^2}{S_{XX} \cdot S_{YY}} \tag{3.39}$$

No caso da regressão simples podemos obter r^2 através da razão entre a Soma dos Quadrados Explicada ($S_{\widehat{Y}Y}$ ou SQE) com a Soma Total dos Quadrados (S_{YY} ou SQT). Quanto menor for a diferença entre média das observações e o valor estimado, maior será o poder explicativo do modelo. Esta diferença é a Soma dos Quadrados dos Resíduos ($S_{\widehat{Y}Y}$ ou SQR) [38].

$$r^2 = \frac{\sum_{i=1}^{n}(\widehat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = \frac{S_{\widehat{Y}Y}}{S_{YY}} \tag{3.40}$$

ou

$$r^2 = \frac{SQE}{SQT} = 1 - \frac{SQR}{SQT} \tag{3.41}$$

onde $\widehat{Y}$ é o valor estimado (previsão) que pode ser calculado de nosso modelo.

3.9.3 *Coeficiente de Determinação Ajustado*

O valor do coeficiente de determinação ajustado $\bar{r}^2$ é uma alternativa a mais para os estatísticos, sendo mais utilizada no caso da regressão linear múltipla. No caso da regressão linear simples temos

$$\bar{r}^2 = 1 - \left(\frac{n-1}{n-p}\right) \cdot \left(1 - r^2\right) \tag{3.42}$$

onde $p = k + 1$ representa o número de variáveis explicativas mais a constante.

3.10 INFERÊNCIA ESTATÍSTICA

A Inferência estatística é um conjunto de técnicas que objetiva estudar uma população através de evidências fornecidas por uma amostra retirada desta população.

3.10.1 *Teste de Hipóteses e Intervalo de Confiança para β_0*

Vamos testar a hipótese de que o β_0 é igual a um determinado valor, denotado por β_{00}. Desta forma, sejam as hipóteses

$$\begin{cases} H_0 : \beta_{00} = 0 \\ H_1 : \beta_{00} \neq 0 \end{cases}$$

Então, da variância (equação 3.22) e do desvio padrão (equação 3.24) de β_0, respectivamente

$$Var(\widehat{\beta}_0) = \sigma^2_{\widehat{\beta}_0} = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}\right) \qquad e \qquad \sigma_{\widehat{\beta}_0} = \sqrt{\sigma^2_{\widehat{\beta}_0}}$$

teremos

$$\widehat{\beta}_0 \sim N\left(\beta_0,\ \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}\right)\right)$$

Assim, sob H_0 temos que

$$N_0 = \frac{\widehat{\beta}_0 - \beta_{00}}{\sqrt{Var(\widehat{\beta}_0)}} \sim N(0,1)$$

Além disso, seja

$$\chi = \frac{(n-2)QMR}{\sigma^2} \sim \chi^2_{(n-2)}$$

Como as variáveis aleatórias N_0 e χ são independentes, segue que

$$T = \frac{N_0}{\sqrt{\dfrac{\chi}{n-2}}} = \frac{\dfrac{\widehat{\beta}_0 - \beta_{00}}{\sqrt{\sigma^2\left(\dfrac{1}{n} + \dfrac{\bar{X}^2}{S_{XX}}\right)}}}{\sqrt{\dfrac{\dfrac{(n-2)QMR}{\sigma^2}}{n-2}}} = \frac{\widehat{\beta}_0 - \beta_{00}}{\sigma_{\widehat{\beta_0}}} \sim t_{(n-2)},$$

ou seja, T tem distribuição t de Student (Figura 33 no Anexo) com n-2 graus de liberdade. Logo, intervalos de confiança e testes a respeito de β_0 podem ser realizados utilizando a distribuição t.

Portanto, para testar as hipóteses do modelo teremos

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases}$$

Assim, a estatística do teste é dada por

$$T_0 = \frac{\widehat{\beta}_0}{\sigma_{\widehat{\beta_0}}} \sim t_{(n-2)} \tag{3.43}$$

Logo, rejeitamos H_0 com um nível de confiança de $(1-\alpha)100\%$ se $\mid T_0 \mid > t_{(1-\alpha/2, n-2)}$.

Geralmente adotamos $\alpha = 0,05 = 5\%$.

O intervalo de confiança para β_0 com $(1-\alpha)\cdot 100\%$ é dado por

$$\left[\widehat{\beta}_0 - t_{\left(1-\frac{\alpha}{2}; n-2\right)} \cdot \sigma_{\widehat{\beta_0}} \quad ; \quad \widehat{\beta}_0 + t_{\left(1-\frac{\alpha}{2}; n-2\right)} \cdot \sigma_{\widehat{\beta_0}}\right] \tag{3.44}$$

3.10.2 *Teste de Hipóteses e Intervalo de Confiança para* β_1

Inferência sobre β_1 é mais frequente já que por meio deste parâmetro temos um indicativo da existência ou não de associação linear entre as variáveis envolvidas.

Similarmente ao parâmetro β_0, consideremos as hipóteses

$$\begin{cases} H_0 : \beta_{10} = 0 \\ H_1 : \beta_{10} \neq 0 \end{cases}$$

Então, da variância (equação 3.23) e do desvio padrão (equação 3.25) de β_1, respectivamente

$$var(\widehat{\beta_1}) = \sigma^2_{\widehat{\beta_1}} = \frac{\sigma^2}{S_{XX}} \qquad e \qquad \sigma_{\widehat{\beta_1}} = \sqrt{\sigma^2_{\widehat{\beta_1}}}$$

teremos

$$\widehat{\beta_1} \sim N\left(\beta_1, \frac{\sigma^2}{S_{XX}}\right)$$

Assim, sob H_1 temos que

$$N_1 = \frac{\widehat{\beta_1} - \beta_{10}}{\sqrt{Var(\widehat{\beta_1})}} \sim N(0, 1)$$

Além disso, seja

$$\chi = \frac{(n-2)QMR}{\sigma^2} \sim \chi^2_{(n-2)}.$$

Como as variáveis aleatórias N_1 e χ são independentes, segue que

$$T = \frac{N_1}{\sqrt{\dfrac{\chi}{n-2}}} = \frac{\dfrac{\widehat{\beta_1} - \beta_{10}}{\sqrt{\dfrac{\sigma^2}{S_{XX}}}}}{\sqrt{\dfrac{\dfrac{(n-2)QMR}{\sigma^2}}{n-2}}} = \frac{\widehat{\beta_1} - \beta_{10}}{\sigma_{\widehat{\beta_1}}} \sim t_{(n-2)},$$

ou seja, T tem distribuição t de Student (Figura 33 com n-2 graus de liberdade. Logo, intervalos de confiança e testes a respeito de β_1 podem ser realizados utilizando a distribuição t.

Portanto, para testar as hipóteses do modelo teremos

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

Assim, a estatística do teste é dada por

$$T_0 = \frac{\widehat{\beta_1}}{\sigma_{\widehat{\beta_1}}} \sim t_{(n-2)} \tag{3.45}$$

Logo, rejeitamos H_0 com um nível de confiança de $(1-\alpha)100\%$ se $\mid T_0 \mid > t_{(1-\alpha/2,n-2)}$.

O intervalo de confiança para β_1 com $(1-\alpha)\cdot 100\%$ é dado por

$$\left[\widehat{\beta}_1 - t_{\left(1-\frac{\alpha}{2};n-2\right)} \cdot \sigma_{\widehat{\beta_1}} \quad ; \quad \widehat{\beta}_1 + t_{\left(1-\frac{\alpha}{2};n-2\right)} \cdot \sigma_{\widehat{\beta_1}}\right] \tag{3.46}$$

3.10.3 *Significância do coeficiente de correlação*

Para comprovarmos se o coeficiente de correlação é significativo, devemos realizar o seguinte teste de hipóteses:

Hipóteses:
$H_0 : r = 0$
$H_1 : r \neq 0$

Então, devemos usar a expressão de t de Student.

$$t_c = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{3.47}$$

O valor crítico da estatística t de Student é obtido definindo-se $n-2$. Caso o valor de t_c seja superior ao valor crítico de t, devermos rejeitar a hipótese nula. Se a hipótese nula, ao nível de significância β_0 , for rejeitada podemos concluir que efetivamente existe uma relação significativa entre as variáveis.

3.10.4 *Nosso modelo é significativo?*

Devemos sempre fazer a pergunta se nosso modelo é significativo?

Então devemos observar os valores:

I - O valor do coeficiente de determinação ajustado ($\bar{r}^2$) é maior que 0,70 = 70%?
No nosso exemplo o valor de $r^2 = 0,998865242 = 99,87\% > 70\%$, portanto tivemos um bom ajuste da reta.

II - Quanto vale o erro?
Observamos o P-valor temos: $p = 1,56 \cdot 10^{-20}$ que em porcentagem representa

um erro muito pequeno, muito menor que 5% = $0,05$, pois nosso nível de significância foi de 95%, portanto os coeficientes são significativos.

III - Devemos verificar se o valor de $\widehat{\beta}$ no intervalo de confiança é diferente de zero. Observando o intervalo de confiança [0,702395403 , 0,731350587], errando no máximo 5%, observamos que $\widehat{\beta_1}$ não assumi o valor zero.

Podemos então concluir que nosso MRLS é significativo.

3.11 EXEMPLO DE MRLS

Retornando ao exemplo do capítulo 2 sobre despesas de consumo pessoal e produto interno bruto dos EUA no período 2000-2014, em bilhões de dólares (Tabela 4).

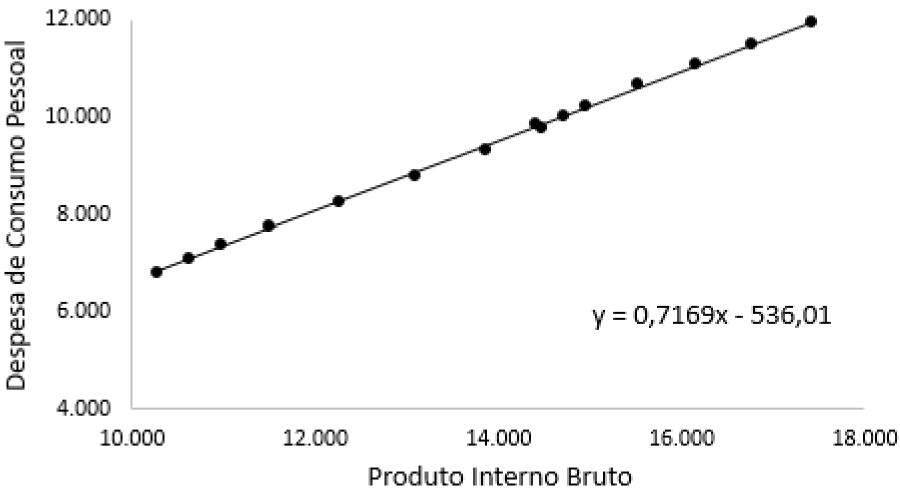
Utilizaremos os programas: Action Stat e o Excel para construir uma planilha (Tabela 11) e gerar os resultados para conferência. Reconstruímos o gráfico (Figura 11) com a linha de tendência [3] e a equação de regressão linear (Figura 19). Também poderíamos utilizar como ferramentas: o Stata 12.0, o Gretl e a calculadora científica, para auxiliar na resolução dos problemas de regressão linear.

3 Uma linha de tendência linear é uma linha reta de melhor ajuste usada com conjuntos de dados lineares simples. Seus dados serão lineares se o padrão nos pontos de dados se parecer com uma linha. Uma linha de tendência linear geralmente mostra que algo está aumentando ou diminuindo com uma taxa fixa.

Tabela 11: Consumo Pessoal e PIB dos EUA no período 2000-2014, em bilhões de dólares

Ano	PIB (X)	DCP (Y)	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
2000	10.284,8	6.792,4	-3.519,8	-2.567,7	9.037.728,34	12.388.757,39	6.593.117,53
2001	10.621,8	7.103,1	-3.182,8	-2.257,0	7.183.525,59	10.130.003,65	5.094.079,09
2002	10.977,5	7.384,1	-2.827,1	-1.976,0	5.586.302,58	7.992.305,94	3.904.602,35
2003	11.510,7	7.765,5	-2.293,9	-1.594,6	3.657.815,08	5.261.824,28	2.542.770,42
2004	12.274,9	8.260,0	-1.529,7	-1.100,1	1.682.796,50	2339.880,11	1.210.234,68
2005	13.093,7	8.794,1	-710,9	-566,0	402.355,27	505.331,42	320.363,55
2006	13.855,9	9.304,0	51,3	-56,1	-2.880,14	2.635,11	3.147,96
2007	14.477,6	9.750,5	673,0	390,4	262.747,73	452.973,87	152.406,95
2008	14.718,6	10.013,6	914,0	653,5	597.314,69	835.456,93	427.053,54
2009	14.418,7	9.847,0	614,1	486,9	299.017,43	377.159,75	237.065,12
2010	14.964,4	10.202,2	1.159,8	842,1	976.687,92	1.345.213,36	709.121,18
2011	15.517,9	10.689,3	1.713,3	1.329,2	2.277.351,24	2.935.511,11	1.766.754,92
2012	16.163,2	11.083,1	2.358,6	1.723,0	4.063.909,51	5.563.151,20	2.968.706,03
2013	16.768,1	11.484,3	2.963,5	2.124,2	6.295.117,75	8.782.529,82	4.512.197,32
2014	17.420,7	11.928,4	3.616,1	2.568,3	9.287.291,13	13.076.420,28	6.596.130,65
Totais	207.068,5	140.401,6	0,0	0,0	51.607.080,60	71.989.154,23	37.037.751,27
Médias	13.804,6	9.360,1					

Figura 19: Gráfico DCP x PIB dos EUA com os parâmetros



3.11.1 *Resultados obtidos dos programas*

Com o auxílio do Excel (Figura 20) e do Action Stat (Figura 21) obtemos os seguintes resultados do nosso exemplo.

Figura 20: Resultados obtidos dos dados pelo Excel

Estatística de regressão	
R múltiplo	0,999432461
R-Quadrado	0,998865243
R-Quadrado Ajustado	0,998777954
Erro Padrão	56,85936335
Observações	15

ANOVA

	gl	SQ	MQ	F	F de significação
Regressão	1	36995722,44	36995722,4	11443,2	1,56209E-20
Resíduo	13	42028,83361	3232,9872		
Total	14	37037751,27			

	Coeficientes	Erro Padrão	t Stat	valor-P	95% Inferiores	95% Superiores
Intercept	-536,014384	93,66820835	-5,72247931	7,02E-05	-738,3722459	-333,656523
X	0,716872995	0,006701445	106,972899	1,56E-20	0,702395403	0,731350587

Figura 21: Resultados obtidos dos dados pelo Action Stat

Tabela da ANOVA					
Fatores	G.L.	Soma de Quadrados	Quadrado Médio	Estat. F	P-valor
PIB__X_	1	36995722,44	36995722,44	11443,201	1,5621E-20
Resíduos	13	42028,83361	3232,987201		

Análise exploratória (resíduos)					
Mínimo	1Q	Mediana	Média	3Q	Máximo
-92,91	-44,48	-0,2837	4,395E-18	46,64	101

Coeficientes				
Preditor	Estimativa	Desvio Padrão	Estat.t	P-valor
Intercepto	-536,0143845	93,66820835	-5,722479312	7,0244E-05
PIB__X_	0,716872995	0,006701445	106,9728985	1,5621E-20

Medida Descritiva da Qualidade do Ajuste			
Desvio Padrão dos Resíduos	Graus de Liberdade	R^2	R^2 Ajustado
56,85936335	13	0,998865243	0,998777954

Intervalo de confiança para os parâmetros		
	0,025	0,975
(Intercept)	-738,3722459	-333,656523
PIB__X_	0,702395403	0,731350587

3.11.2 *Os estimadores $\widehat{\beta}_0$ e $\widehat{\beta}_1$*

Vamos determinar os valores dos estimadores $\widehat{\beta}_0$ e $\widehat{\beta}_1$ utilizando as equações (3.9) e (3.13), receptivamente.

Os valores poderão ter uma pequena diferença em virtude da quantidade de casas fixadas para a resolução dos exercícios.

Substituindo os valores constante na Tabela 11 e substituindo na equação (3.13) para determinar o valor do estimador $\hat{\beta_1}$ teremos:

$$\widehat{\beta_1} = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{i=1}^{i}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{i}(X_i - \bar{X})^2}$$

$$\widehat{\beta_1} = \frac{51.607.080,60}{71.898.154,23}$$

$$\widehat{\beta_1} = 0,716873$$

Agora substituindo os dados na equação (3.9) e utilizando o valor encontrado para $\widehat{\beta_1}$ podemos determinar o valor do $\widehat{\beta_0}$.

$$\widehat{\beta_0} = \bar{Y} - \widehat{\beta_1}\bar{X}$$

$$\widehat{\beta_0} = 9360,1 - 0,716873 \cdot 13804,6$$

$$\widehat{\beta_0} = -536,0144$$

3.11.3 *O Estimador Quadrado Médio dos Resíduos (QMR)*

Utilizaremos a equação (3.15) para determinar o estimador, onde o n é o número de observações (neste caso $n = 15$) e o grau de liberdade [4] (gl) é dado por $gl = n - 2 = 15 - 2 = 13$.

$$\widehat{\sigma}^2 = \frac{S_{YY} - \widehat{\beta_1} \cdot S_{XY}}{n - 2}$$

$$\widehat{\sigma}^2 = \frac{37.037.751,27 - 0,716873 \cdot 51.607.080,60}{15 - 2}$$

$$QMR = \widehat{\sigma}^2 = 3.232,9872$$

3.11.4 *O Desvio Padrão dos Resíduos*

Utilizando a equação (3.16) teremos

$$\widehat{\sigma} = \sqrt{\widehat{\sigma}^2}$$

4 Grau de liberdade é, em estatística, o número de determinações independentes (dimensão da amostra) menos o número de parâmetros estatísticos a serem avaliados na população.

$$\widehat{\sigma} = \sqrt{3.232,9872}$$
$$\widehat{\sigma} = 56,85936335$$

3.11.5 *A Variância e Desvio Padrão de $\widehat{\beta_0}$ e $\widehat{\beta_1}$.*

Para determinarmos a variância dos estimadores $\widehat{\beta_0}$ e $\widehat{\beta_1}$ utilizaremos as equações (3.22) e (3.23), respectivamente, então

$$var(\widehat{\beta_0}) = \sigma^2_{\widehat{\beta_0}} = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 = \frac{\sum X_i^2}{nS_{XX}}\sigma^2 = \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}\right)$$

$$var(\widehat{\beta_0}) = \sigma^2_{\widehat{\beta_0}} = 3.232,9872 \cdot \left(\frac{1}{15} + \frac{(13.804,56667)^2}{71.989.154,23}\right)$$

$$var(\widehat{\beta_0}) = \sigma^2_{\widehat{\beta_0}} = 8.773,733256$$

$$var(\widehat{\beta_1}) = \sigma^2_{\widehat{\beta_1}} = \frac{\sigma^2}{\sum x_i^2} = \frac{\sigma^2}{S_{XX}}$$

$$var(\widehat{\beta_1}) = \sigma^2_{\widehat{\beta_1}} = \frac{3.232,9872}{71.989.154,23}$$

$$var(\widehat{\beta_1}) = \sigma^2_{\widehat{\beta_1}} = 4,49094 \cdot 10^{-5}$$

O desvio padrão dos estimadores $\widehat{\beta_0}$ e $\widehat{\beta_1}$, é calculado com as equações (3.24) e (3.25), respectivamente

$$\sigma_{\widehat{\beta_0}} = \sqrt{\sigma^2_{\widehat{\beta_0}}}$$
$$\sigma_{\widehat{\beta_0}} = \sqrt{8.773,733256}$$
$$\sigma_{\widehat{\beta_0}} = 93,66820835$$

$$\sigma_{\widehat{\beta_1}} = \sqrt{\sigma^2_{\widehat{\beta_1}}}$$
$$\sigma_{\widehat{\beta_1}} = \sqrt{4,49094 \cdot 10^{-5}}$$
$$\sigma_{\widehat{\beta_1}} = 0,006701445$$

3.11.6 *Tabela da ANOVA para o exemplo*

Já calculado anteriormente.

$$QMR = 3.232,9872$$
$$\widehat{\beta_1} = 0,716873$$

Dados extraídos da tabela 11, temos:

$$SQT = S_{YY} = 37.037.751,27$$
$$S_{XX} = \sum_{i=1}^{n}(X_i - \bar{X}) = 71.989.154,23$$
$$\sum_{i=1}^{n}(Y_i) = 140.401,6$$

Calculando os demais

$$SQR = 37.037.751,27 - 0,716873 \cdot 51.607.080,60$$
$$SQR = 42.028,83$$

$$SQE = SQT - SQR$$
$$SQE = 37.037.751,27 - 42.028,83$$
$$SQE = QME = 36.995.722,44$$

Tabela 12: Tabela ANOVA referente ao exemplo

Fonte	**gl**	**Soma de Quadrados**	**Quadrado Médio**
Regressão	1	$SQE = 36.995.722,44$	$QME = 36.995.722,44$
Resíduo	13	$SQR = 42.028,83$	$QMR = 3.232,9872$
Total	14	$SQT = 37.037.751,27$	-

3.11.7 *O Teste F para um nível de confiança de 95%.*

Temos que testar as hipóteses.

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

Vamos determinar o F_0

$$F_0 = \frac{QME}{QMR} = \frac{36.995.722,44}{3.232,9872}$$
$$F_0 = 11.443,20102$$

Utilizando a tabela t (Figura 33) para $\alpha = 0,05 = 5\%$, obtemos que $F_{(0,95;1;13)} = (2,1604)^2 = 4,6673$

Logo,

$$F_0 = 11.443,20102 > 4,6673 = F_{(0,95;1;13)}$$

Portanto, rejeitamos H_0 com um nível de confiança de 95% e concluímos que a variável explicativa tem correlação com a variável resposta.

3.11.8 *O Coeficiente de Correlação e de Determinação*

O Coeficiente de Correlação de Pearson

Agora vamos determinar o coeficiente de correlação e classifica-lo de acordo com a Tabela 10, portanto vamos substituir os dados na equação 3.38

$$r = \frac{S_{XY}}{\sqrt{S_{XX} \cdot S_{YY}}}$$

$$r = \frac{51.607.080,6}{\sqrt{71.989.154,23 \cdot 37.037.751,27}}$$

$$r = 0,99943246$$

O resultado indica que existe uma correlação positiva muito forte entre as variáveis.

O Coeficiente de Determinação

Em nosso exemplo anterior, o valor do coeficiente de correlação de Pearson (r) encontrado foi de $r = 0,99943246$, portanto o coeficiente de determinação (r^2) será

$$r^2 = 0,99943246^2$$
$$r^2 = 0,998865242 = 99,87\%$$

portanto, a qualidade do ajuste obtido do nosso modelo foi de $99,87\%$.

O Coeficiente de Determinação Ajustado

Utilizaremos a equação 3.42 para determinar o coeficiente de determinação ajustado, então

$$\bar{r}^2 = 1 - \left(\frac{n-1}{n-2}\right) \cdot \left(1 - r^2\right)$$

$$\bar{r}^2 = 1 - \left(\frac{15-1}{15-2}\right) \cdot (1 - 0,998865)$$

$$\bar{r}^2 = 0,998777954$$

3.11.9 *Os testes de hipóteses t e os intervalos de confiança para os parâmetros β_0 e β_1*

Já calculamos anteriormente: $\widehat{\beta_0} = -536,0144$, $\widehat{\beta_1} = 0,716873$, $\sigma_{\widehat{\beta_0}} = 93,66820835$ e $\sigma_{\widehat{\beta_1}} = 0,006701445$.

Para β_0, devemos testar as hipóteses

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases}$$

A estatística do teste é dado pela equação 3.45

$$T_0 = \frac{\widehat{\beta_0}}{\sigma_{\widehat{\beta_0}}} = \frac{-536,0144}{93,66820835} = -5,722479$$

Teremos uma distribuição normal t com um nível de significância (α) de 5% e grau de liberdade de 13 e consultando a Tabela 33 do Anexo, temos $t = 2,1604$. Como $T_0 = -5,722479$

O intervalo de confiança de 95% para β_0 é dado pela equação 3.46

$$\begin{aligned}
[\widehat{\beta_0} - (t \cdot \sigma_{\widehat{\beta_0}}) \quad &; \quad \widehat{\beta_0} + (t \cdot \sigma_{\widehat{\beta_0}})] \\
[-536,0144 - (2,1604 \cdot 93,66820835) \quad &; \quad -536,0144 + (2,1604 \cdot 93,66820835)] \\
[-738,3752 \quad &; \quad -333,6536]
\end{aligned}$$

Para β_1, devemos testar as hipóteses

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

A estatística do teste, sob H_0 é dada por

$$T_0 = \frac{\widehat{\beta_1}}{\sigma_{\widehat{\beta_1}}} = \frac{0,716873}{0,006701445} = 106,972899$$

O intervalo de confiança de 95% para β_1 é dado por

$$\begin{aligned}
[\widehat{\beta_1} - (t \cdot \sigma_{\widehat{\beta_1}}) \quad &; \quad \widehat{\beta_1} + (t \cdot \sigma_{\widehat{\beta_1}})] \\
[0,716873 - (2,1604 \cdot 0,006701445) \quad &; \quad 0,716873 + (2,1604 \cdot 0,006701445)] \\
[0,702395 \quad &; \quad 0,731350]
\end{aligned}$$

3.12 EXERCÍCIO RESOLVIDO DE MRLS

1- O gerente administrativo de uma cadeia de hipermercados deseja desenvolver um modelo com a finalidade de estimar as vendas semanais (em milhares de reais).

- Y - vendas semanais.
- X - número de clientes.

Tabela 13: Número de clientes e vendas semanais de uma rede de hipermercado

Lojas	Clientes (X)	Vendas (Y)	Lojas	Clientes (X)	Vendas (Y)
1	907	11,20	11	679	7,63
2	926	11,05	12	872	9,43
3	506	6,84	13	924	9,46
4	741	9,21	14	607	7,64
5	789	9,42	15	452	6,92
6	889	10,08	16	729	8,95
7	874	9,45	17	794	9,33
8	510	6,73	18	844	10,23
9	529	7,24	19	1010	11,77
10	420	6,12	20	621	7,41

Determine:

a-) os estimadores S_{XX}, S_{XY} e S_{YY}.

b-) os estimadores dos parâmetros $\widehat{\beta_0}$ e $\widehat{\beta_1}$.

c-) a equação de regressão linear simples.

d-) a soma dos quadrados dos resíduos (SQR).

e-) o quadrado médio dos resíduos (QMR).

f-) o desvio padrão dos resíduos.

g-) a variância de $\widehat{\beta_0}$ e $\widehat{\beta_1}$.

h-) o desvio padrão de $\widehat{\beta_0}$ e $\widehat{\beta_1}$.

i-) a análise de variância (ANOVA) completando a tabela.

j-) o teste F para um nível de confiança de 95%.

k-) o coeficiente de correlação de Pearson.

l-) o coeficiente de determinação.

m-) o coeficiente de determinação ajustado.

n-) os testes de hipóteses t e o intervalo de confiança de 95% para os parâmetros β_0 e β_1 .

o-) o intervalo de confiança de 95%.

Resolução:

a-) Vamos construir a tabela.

Tabela 14: Número de clientes e vendas semanais, em milhares de reais

Lojas	**Clientes (X)**	**Vendas (Y)**	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
01	907	11,20	421,07	30.923,22	5,73
02	926	11,05	437,34	37.966,52	5,04
03	506	6,84	442,53	50.692,52	3,86
04	741	9,21	3,98	97,02	0,16
05	789	9,42	35,55	3.346,62	0,38
06	889	10,08	201,18	24.916,62	1,62
07	874	9,45	92,07	20.406,12	0,42
08	510	6,73	459,00	48.907,32	4,31
09	529	7,24	316,47	40.864,62	2,45
10	420	6,12	835,59	96.814,32	7,21
11	679	7,63	61,30	2.719,62	1,38
12	872	9,43	87,96	19.838,72	0,39
13	924	9,46	126,22	37.191,12	0,43
14	607	7,64	144,70	15.413,22	1,36
15	452	6,92	526,,34	77.924,72	3,56
16	729	8,95	-0,31	4,62	0,02
17	794	9,33	32,96	3.950,12	0,28
18	844	10,23	160,75	12.735,12	2,03
19	1.010	11,77	826,65	77.757,32	8,79
20	621	7,41	153,71	12.133,02	1,95
Totais	14.623	176,11	5.365,07	614.602,55	51,3605
Médias	731,15	8,8055			

$$S_{XX} = \sum_{i=1}^{n}(X_i - \bar{X})^2 = 614.602,55$$

$$S_{XY} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = 5.365,07$$

$$S_{YY} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = 51,3605$$

b-) $\widehat{\beta}_1 = \dfrac{S_{XY}}{S_{XX}} = \dfrac{5.365,07}{614.602,55} = 0,00873$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1\bar{X} = 8,8055 - (0,00873)(731,15) = 2,423$$

c-) $\widehat{Y} = 2,423 + 0,00873X$

d-) $SQR = S_{YY} - \widehat{\beta}_1 \cdot S_{XY} = 51,3605 - (0,00873)(5.365,07) = 4,52695$

e-) $QMR = \widehat{\sigma}^2 = \dfrac{SQR}{n-2} = \dfrac{4,52695}{20-2} = 0,251497$

f-) $\widehat{\sigma} = \sqrt{\widehat{\sigma}^2} = \sqrt{0,251497} = 0,501495$

g-) $var(\widehat{\beta}_0) = \sigma^2_{\widehat{\beta}_0} = \sigma^2\left(\dfrac{1}{n} + \dfrac{\bar{X}^2}{S_{XX}}\right) = 0,251497 \cdot \left(\dfrac{1}{20} + \dfrac{(731,15)^2}{614.602,55}\right) = 0,231327$

$var(\widehat{\beta}_1) = \sigma^2_{\widehat{\beta}_1} = \dfrac{\sigma^2}{S_{XX}} = \dfrac{0,251497}{614.602,55} = 4,092 \cdot 10^{-7}$

h-) $\sigma_{\widehat{\beta}_0} = \sqrt{\sigma^2_{\widehat{\alpha}}} = \sqrt{0,231327} = 0,480964$

$\sigma_{\widehat{\beta}_1} = \sqrt{\sigma^2_{\widehat{\beta}}} = \sqrt{4,092 \cdot 10^{-7}} = 0,0006397$

i-) $SQT = S_{YY} = 51,3605$

$SQE = QME = SQT - SQR = 51,3605 - 4,52335 = 46,8335$

Tabela 15: Tabela ANOVA do exercício 1

Fonte	**gl**	**Soma de Quadrados**	**Quadrado Médio**
Regressão	1	$SQE = 46,8335$	$QME = 46,8335$
Resíduo	20-2=18	$SQR = 4,52335$	$QMR = 0,251497$
Total	20-1=19	$SQT = 51,3605$	

j-) Temos que testar as hipóteses.

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

Vamos determinar o F_0

$$F_0 = \frac{QME}{QMR} = \frac{46,8335}{0,251497} = 186,21875$$

Utilizando a tabela t (Figura 33) para $\alpha = 0,05 = 5\%$, obtemos que $F_{(0,95;1;13)} = (2,1604)^2 = 4,6673$

Logo, $F_0 = 186,38 > 4,6673 = F_{(0,95;1;13)}$
Portanto, rejeitamos H_0 com um nível de confiança de 95% e concluímos que a variável explicativa tem correlação com a variável resposta.

k-) $r = \dfrac{S_{XY}}{\sqrt{S_{XX} \cdot S_{YY}}} = \dfrac{5.365,07}{\sqrt{614.602,55 \cdot 51,3605}} = 0,954913 = 91,5\%.$

Portanto, possui uma correlação muito forte.

l-) $r^2 = 0,954913^2 = 0,911859$
Portanto, a qualidade do ajuste obtido do nosso modelo foi de $91,2\%$.

m-) $\bar{r}^2 = 1 - \left(\dfrac{n-1}{n-2}\right) \cdot \left(1 - r^2\right) = 1 - \left(\dfrac{20-1}{20-2}\right) \cdot (1 - 0,911859) = 0,906962$

n-) Para β_0, devemos testar as hipóteses
$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases}$$
A estatística do teste é dado pela equação 3.45

$$T_0 = \frac{\widehat{\beta_0}}{\sigma_{\widehat{\beta_0}}} = \frac{2,423}{0,480964} = 5,0378$$

Teremos uma distribuição normal t com um nível de significância (α) de 5% e grau de liberdade de 18 e consultando a Tabela 33 do Anexo, temos $t = 2,1009$. Como $T_0 = 5,0378$

O intervalo de confiança de 95% para β_0 é dado pela equação 3.46

$$[\widehat{\beta_0} - (t \cdot \sigma_{\widehat{\beta_0}}) \quad ; \quad \widehat{\beta_0} + (t \cdot \sigma_{\widehat{\beta_0}})]$$
$$[2,423 - (2,1009 \cdot 0,4809642) \quad ; \quad 2,423 + (2,1009 \cdot 0,4809642)]$$
$$[1,4125 \quad ; \quad 3,4335]$$

Para β_1, devemos testar as hipóteses

$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$ A estatística do teste, sob H_0 é dada por

$$T_0 = \frac{\widehat{\beta_1}}{\sigma_{\widehat{\beta_1}}} = \frac{0,00873}{0,00063969} = 13,647$$

O intervalo de confiança de 95% para β_1 é dado por

$$[\widehat{\beta_1} - (t \cdot \sigma_{\widehat{\beta_1}}) \quad ; \quad \widehat{\beta_1} + (t \cdot \sigma_{\widehat{\beta_1}})]$$
$$[0,00873 - (2,1009 \cdot 0,00063969) \quad ; \quad 0,00873 + (2,1009 \cdot 0,00063969)]$$
$$[0,007386 \quad ; \quad 0,010074]$$

3.13 EXERCÍCIOS PROPOSTOS DE MRLS

1-) Uma empresa resolveu estudar a variação da demanda (em unidades) de seu produto em função do preço de venda (em reais) praticado. Para isso, foram coletados os seguintes dados:

Tabela 16: Variação da demanda do produto em função do preço de venda.

Preço (X)	36	43	49	55	61	63	69	72	74	77
Demanda (Y)	350	330	296	252	230	218	203	196	188	167

Determine:

a-) os estimadores S_{XX}, S_{XY} e S_{YY}.

b-) os estimadores dos parâmetros $\widehat{\beta_0}$ e $\widehat{\beta_1}$.

c-) a equação de regressão linear simples.

d-) a soma dos quadrados dos resíduos (SQR).

e-) o quadrado médio dos resíduos (QMR).

f-) o desvio padrão dos resíduos.

g-) a variância de $\widehat{\beta_0}$ e $\widehat{\beta_1}$.

h-) o desvio padrão de $\widehat{\beta_0}$ e $\widehat{\beta_1}$.

i-) a análise de variância (ANOVA) completando a tabela.

j-) o teste F para um nível de confiança de 95%.

k-) o coeficiente de correlação de Pearson.

l-) o coeficiente de determinação.

m-) o coeficiente de determinação ajustado.

n-) os testes de hipóteses t e o intervalo de confiança de 95% para os parâmetros β_0 e β_1 .

o-) o intervalo de confiança de 95%.

2-) O coordenador de um curso de graduação de Matemática com a intenção de verificar se existe alguma correlação entre as notas de Matemática e de Estatística de uma turma sorteou 15 alunos aleatoriamente e construiu uma tabela com as respectivas notas. Utilize as perguntas da questão número 1 e verifique se existe esta correlação.

Tabela 17: Notas de Matemática e Estatística.

Mat. (X)	5	8	7	10	6	7	9	3	8	2	7	9	4	6	3
Est. (Y)	6	8	9	10	5	7	8	4	6	2	6	7	5	7	4

Determine:

a-) os estimadores S_{XX}, S_{XY} e S_{YY}.

b-) os estimadores dos parâmetros $\widehat{\beta_0}$ e $\widehat{\beta_1}$.

c-) a equação de regressão linear simples.

d-) a soma dos quadrados dos resíduos (SQR).

e-) o quadrado médio dos resíduos (QMR).

f-) o desvio padrão dos resíduos.

g-) a variância de $\widehat{\beta_0}$ e $\widehat{\beta_1}$.

h-) o desvio padrão de $\widehat{\beta_0}$ e $\widehat{\beta_1}$.

i-) a análise de variância (ANOVA) completando a tabela.

j-) o teste F para um nível de confiança de 95%.

k-) o coeficiente de correlação de Pearson.

l-) o coeficiente de determinação.

m-) o coeficiente de determinação ajustado.

n-) os testes de hipóteses t e o intervalo de confiança de 95% para os parâmetros β_0 e β_1 .

o-) o intervalo de confiança de 95%.

4

MODELO DE REGRESSÃO LINEAR MÚLTIPLA

4.1 INTRODUÇÃO

Neste capítulo vamos tratar do modelo de regressão linear múltipla (MRLM) que descreve a relação entre duas ou mais variáveis independentes (ou variáveis explicativas) X, de modo que vamos encontrar frequentemente modelos da forma

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \varepsilon_i \tag{4.1}$$

na qual a variável dependente Y é relacionada às variáveis independentes X_1, X_2 etc., e ao distúrbio ε.

De maneira similar ao MRLS devemos explicar o comportamento destas variáveis, mas devemos atentar pra que elas não sejam correlacionadas entre si, podendo criar uma multicolinearidade [1], o que não é desejável ao modelo de regressão, pois poderiam explicar a mesma coisa e prejudicar a análise do modelo proposto. [36]

Este problema da multicolinearidade entre as variáveis independentes pode ocorrer devido a existência que os valores dos coeficientes angulares (β_1, β_2, ...) associados as variáveis independentes podem estar viesados, portanto não podemos confiar na apuração dos resultados obtidos na simulação do modelo e consequentemente não podemos obter conclusões e realizar os testes de regressão.

Podemos verificar a ocorrência de problemas de multicolinearidade pela obtenção da matriz de correlação entre as variáveis do MRLM.

1 O termo Multicolinearidade é utilizado para indicar a existência forte de correlação entre duas (ou mais) variáveis independentes e a existência de relação linear. O índício mais claro da existência da multicolinearidade é quando o R^2 é bastante alto.

4.2 ESTIMAÇÃO DOS PARÂMETROS DO MRLM PELO MQO

A maioria dos aspectos da análise de regressão linear múltipla podem ser descritos em termos de uma relação entre duas variáveis independentes, sem que a álgebra se torne tão complexa, pelo fato de a inclusão de mais variáveis implicar em fórmulas complicadas ou na utilização álgebra matricial, o que proporciona muito pouco entendimento intuitivo. Vamos supor, daqui por diante, que os dados de Y, X_1 e X_2 tenham sido gerados pelo modelo.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \tag{4.2}$$

Ajustando a linha.

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} \tag{4.3}$$

Aplicando-se o método dos mínimos quadrados, vamos escolher $\widehat{\beta}_0$, $\widehat{\beta}_1$ e $\widehat{\beta}_2$ para minimizar

$$S = \Sigma\varepsilon_i^2 = \Sigma(Y_i - \widehat{Y}_i)^2$$

$$S = \Sigma(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i})^2$$

Derivando parcialmente com relação a $\widehat{\beta}_0$, $\widehat{\beta}_1$ e $\widehat{\beta}_2$, e igualando cada derivada a zero

$$\frac{dS}{d\widehat{\beta}_0} = -2\Sigma(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i}) = 0,$$

$$\frac{dS}{d\widehat{\beta}_1} = -2\Sigma X_{1i}(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i}) = 0,$$

$$\frac{dS}{d\widehat{\beta}_2} = -2\Sigma X_{2i}(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i}) = 0.$$

Num rearranjo, estes dão as três equações normais:

$$\Sigma Y_i = n\widehat{\beta}_0 + \widehat{\beta}_1 \Sigma X_{1i} + \widehat{\beta}_2 \Sigma X_{2i},$$

$$\Sigma Y_i X_{1i} = \widehat{\beta}_0 \Sigma X_{1i} + \widehat{\beta}_1 \Sigma X_{1i}^2 + \widehat{\beta}_2 \Sigma X_{1i} X_{2i},$$

$$\Sigma Y_i X_{2i} = \widehat{\beta}_0 \Sigma X_{2i} + \widehat{\beta}_1 \Sigma X_{1i} X_{2i} + \widehat{\beta}_2 \Sigma X_{2i}^2.$$

Deste modo, temos três equações e três incógnitas $\widehat{\beta}_0$, $\widehat{\beta}_1$ e $\widehat{\beta}_2$ para serem resolvidas em termos da soma dos quadrados e os produtos cruzados obtidos a partir dos dados.

Será reconhecido que as equações normais são facilmente generalizadas para o caso de k variáveis independentes. A j-ésima equação normal será então

$$\Sigma YX_j = \widehat{\beta}_0\Sigma X_j + \widehat{\beta}_1\Sigma X_1X_j + \widehat{\beta}_2\Sigma X_2X_j + \cdots + \widehat{\beta}_j\Sigma X_j^2 + \cdots + \widehat{\beta}_k\Sigma X_kX_j$$

Juntamente com a primeira equação normal

$$\Sigma Y = n\widehat{\beta}_0 + \widehat{\beta}_1\Sigma X_1 + \widehat{\beta}_2\Sigma X_2 + \cdots + \widehat{\beta}_j\Sigma X_j + \cdots + \widehat{\beta}_k\Sigma X_k$$

isso resultará em $(k+1)$ equações normais a resolver para os $(k+1)$ coeficientes.

Alternativamente, podemos adotar o expediente usado no caso da regressão simples, sabendo-se que a primeira equação normal pode ser escrita

$$\bar{Y} = \widehat{\beta}_0 + \widehat{\beta}_1\bar{X}_1 + \widehat{\beta}_2\bar{X}_2$$

e portanto as duas equações restantes serão escritas como (desprezando o subíndice i para maior clareza):

$$\Sigma yx_1 = \widehat{\beta}_1\Sigma x_1^2 + \widehat{\beta}_2\Sigma x_1x_2 \tag{4.4}$$

$$\Sigma yx_2 = \widehat{\beta}_1\Sigma x_1x_2 + \widehat{\beta}_2\Sigma x_2^2 \tag{4.5}$$

onde as letras minúsculas para as variáveis denotam desvios das médias das mesmas.

$$\begin{cases} y = Y_i - \bar{Y} \\ x_1 = X_1 - \bar{X}_1 \\ x_2 = X_2 - \bar{X}_2 \\ y^2 = (Y_i - \bar{Y})^2 \\ x_1^2 = (X_1 - \bar{X}_1)^2 \\ x_2^2 = (X_2 - \bar{X}_2)^2 \\ yx_1 = (Y_i - \bar{Y})(X_1 - \bar{X}_1) \\ yx_2 = (Y_i - \bar{Y})(X_2 - \bar{X}_2) \\ x_1x_2 = (X_1 - \bar{X}_1)(X_2 - \bar{X}_2) \end{cases}$$

Novamente, estas podem ser resolvidas simultaneamente para $\widehat{\beta}_1$ e $\widehat{\beta}_2$, $\widehat{\beta}_0$ sendo obtido como

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1\bar{X}_1 - \widehat{\beta}_2\bar{X}_2 \tag{4.6}$$

ou ainda poderão ser resolvidas diretamente, pois da equação (4.5)

$$\widehat{\beta}_2 = \frac{\Sigma yx_2 - \widehat{\beta}_1\Sigma x_1x_2}{\Sigma x_2^2}$$

e em substituição na equação (4.4):

$$\Sigma yx_1 = \widehat{\beta}_1 \Sigma x_1^2 + \frac{(\Sigma yx_2 - \widehat{\beta}_1 \Sigma x_1 x_2)}{\Sigma x_2^2} \Sigma x_1 x_2$$

ou

$$\widehat{\beta}_1 = \frac{\Sigma yx_1 \Sigma x_2^2 - \Sigma yx_2 \Sigma x_1 x_2}{\Sigma x_1^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2} \tag{4.7}$$

Igualmente,

$$\widehat{\beta}_2 = \frac{\Sigma yx_2 \Sigma x_1^2 - \Sigma yx_1 \Sigma x_1 x_2}{\Sigma x_1^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2} \tag{4.8}$$

Não iremos reescrever as equações (4.7) e (4.8) com as letras maiúsculas em virtude de ficar muito extensa e desnecessário.

4.2.1 *Exemplo 1 - Regressão Múltipla para duas variáveis independentes*

Exemplo 1 - Para ilustrarmos os cálculos apresentados nas equações (4.6), (4.7) e (4.8) temos a tabela 18 (fictícia) com dados de renda anual (em R$ mil)(Y_i), patrimônio (em R$ mil)(X_1) e a idade (X_2).

Tabela 18: Renda anual em relação ao patrimônio e a idade

i	Y_i	X_1	X_2	y^2	x_1^2	x_2^2	yx_1	yx_2	x_1x_2
1	28	0	23	2.342,56	5.164,82	94,74	3.478,35	471,09	699,50
2	41	4	25	1.253,16	4.605,88	59,8	2.402,48	273,76	524,84
3	48	7	26	806,56	4.207,68	45,34	1.842,21	191,23	436,77
4	54	18	27	501,76	2.901,62	32,87	1.206,61	128,43	308,84
5	66	27	29	108,16	2.013,02	13,94	466,61	38,83	167,50
6	74	42	30	5,76	892,02	7,47	71,68	6,56	81,64
7	86	58	31	92,16	192,28	3,00	-133,12	-16,64	24,04
8	83	66	32	43,56	34,42	0,54	-38,72	-4,84	4,30
9	88	81	33	134,56	83,42	0,07	105,95	3,09	2,44
10	92	99	35	243,36	736,22	5,14	423,28	35,36	61,50
11	92	107	37	243,36	1.234,35	18,20	548,08	66,56	149,90
12	96	115	38	384,16	1.860,48	27,74	845,41	103,23	227,17
13	97	127	39	424,36	3.039,68	39,27	1.135,75	129,09	345,50
14	99	147	42	510,76	5.645,02	85,87	1.698,01	209,43	696,24
15	102	180	44	655,36	11.692,82	126,94	2.768,21	288,43	1.218,30
Soma	1.146	1.078	491	7.749,60	44.303,733	560,933	16.820,8	1.923,6	4.948,467
Média	76,40	71,867	32,733						

Vamos determinar $\widehat{\beta}_1$ utilizando a equação (4.7)

$$\widehat{\beta}_1 = \frac{\Sigma y x_1 \Sigma x_2^2 - \Sigma y x_2 \Sigma x_1 x_2}{\Sigma x_1^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2}$$

$$\widehat{\beta}_1 = \frac{16.820,80 \cdot 560,933 - 1.923,60 \cdot 4.948,467}{44.303,733 \cdot 560,933 - (4.948,467)^2} = \frac{-83.529,3148}{364.100,2128}$$

$$\widehat{\beta}_1 = -0,2294$$

Agora vamos determinar $\widehat{\beta}_2$ utilizando a equação (4.8)

$$\widehat{\beta}_2 = \frac{\Sigma y x_2 \Sigma x_1^2 - \Sigma y x_1 \Sigma x_1 x_2}{\Sigma x_1^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2}$$

$$\widehat{\beta}_1 = \frac{1.923,60 \cdot 44.303,733 - 16.820,80 \cdot 4.948,467}{44.303,733 \cdot 560,933 - (4.948,467)^2} = \frac{1.985.487,085}{364.100,2128}$$

$$\widehat{\beta}_2 = 5,4531$$

Agora vamos determinar $\widehat{\beta}_0$ utilizando a equação (4.6)

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X}_1 - \widehat{\beta}_2 \bar{X}_2$$

$$\widehat{\beta}_0 = 76,40 - (-0,2294) \cdot 71,867 - 5,4531 \cdot 32,733$$

$$\widehat{\beta}_0 = -85,610$$

Portanto, a equação de regressão linear múltipla é

$$\widehat{Y} = -85,610 - 0,2294X_1 + 5,4531X_2$$

Utilizando o Excel para conferir os resultados (figura 22) verificamos que podem ocorrer pequenas diferenças em virtude da quantidade de casas que levamos em consideração nos cálculos, mas são totalmente compatíveis.

		Coeficientes	Erro Padrão	t Stat	valor-P	95% Inferiores	95% Superiores
Intersecção	β_0	-85,6058	83,3183	-1,0275	0,3245	-267,1408	95,9293
Patrim. (R$ mil)	β_1	-0,2294	0,3790	-0,6052	0,5563	-1,0551	0,5964
Idade	β_2	5,4529	3,3682	1,6189	0,1314	-1,8858	12,7916

Figura 22: Análise dos betas do exemplo 1 utilizando o Excel

4.3 ESTIMAÇÃO DOS PARÂMETROS DO MRLM PELA NOTAÇÃO MATRICIAL

Os estimadores de mínimos quadrados dos parâmetros também podem ser encontrados considerando a notação matricial dos dados. Desta forma teremos uma muito compacta. Esta notação matricial é representada como

$$Y = X\beta + \varepsilon$$

com

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{e} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

em que
Y é o vetor $n \times 1$ das observações;
X é uma matriz de dimensão $n \times p$ denominada matriz do modelo com os níveis das variáveis regressoras;
β é um vetor $p \times 1$ com os coeficientes de regressão e
ε é o vetor de dimensão $n \times 1$ com os erros aleatórios.
onde $p = k + 1$ representa o número de variáveis explicativas mais a constante.

Através do método de mínimos quadrados podemos encontrar o vetor $\widehat{\beta}$ que minimiza

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta) =$$

$$= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta = Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

sendo que $Y'X\beta = \beta'X'Y$ pois o produto resulta em um escalar.

A notação X' representa o transposto da matriz X enquanto que Y' e β' representam os transpostos dos vetores Y e β, respectivamente. Usando a técnica de derivação (em termos matriciais) obtemos

$$\frac{\partial L}{\partial \beta} = -2X'Y + 2X'X\beta$$

Igualando a zero e substituindo o vetor β pelo vetor $\widehat{\beta}$, temos

$$(X'X)\widehat{\beta} = X'Y$$

Em geral, a matriz $(X'X)$ é não singular, ou seja, tem determinante diferente de zero, e portanto é invertível. Desta forma, conclui-se que os estimadores para os parâmetros $\beta_j, \quad j = 0, 1, \ldots, k$ são dados pelo vetor

$$\widehat{\beta} = (X'X)^{-1}(X'Y)$$

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_k \end{bmatrix} = [X^t X]^{-1}[X^t Y] \tag{4.9}$$

Portanto, o modelo de regressão linear ajustado e o vetor de resíduos são respectivamente

$$\widehat{Y} = X\widehat{\beta} \quad \text{e} \quad e = Y - \widehat{Y}$$

4.3.1 *Exemplo 2 - Regressão Simples utilizando Notação Matricial*

Inicialmente, vamos a um exemplo utilizando a notação matricial para regressão linear simples.

Exemplo 2 - Com base na tabela 19 vamos determinar os coeficientes $\widehat{\beta}_0$ e $\widehat{\beta}_1$ e a equação de regressão linear simples.

Tabela 19: Quantidade x preço

Ano	**Quantidade (t)**	**Preço (mil R$)**
1	2	4
2	1	6
3	3	3
4	1	5
5	4	1

Temos $k = 1$, pois nosso exemplo contém apenas uma variável independente (o preço) e utilizando a equação 4.9 teremos

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = [X^t X]^{-1}[X^t Y]$$

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \left[\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 6 & 3 & 5 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 4 \\ 1 & 6 \\ 1 & 3 \\ 1 & 5 \\ 1 & 1 \end{bmatrix} \right]^{-1} \times \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 6 & 3 & 5 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ 3 \\ 1 \\ 4 \end{bmatrix}$$

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 5 & 19 \\ 19 & 87 \end{bmatrix}^{-1} \times \begin{bmatrix} 11 \\ 32 \end{bmatrix}$$

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 87/74 & -19/74 \\ -19/74 & 5/74 \end{bmatrix} \times \begin{bmatrix} 11 \\ 32 \end{bmatrix}$$

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 349/74 \\ -49/74 \end{bmatrix}$$

$$\begin{cases} \widehat{\beta}_0 = 4,71622 \\ \widehat{\beta}_1 = 0,66216 \end{cases}$$

Portanto, a equação de regressão linear simples é

$$\widehat{Y} = 4,71622 + 0,66216X$$

Utilizando o Excel para conferir os resultados (figura 23) verificamos que são totalmente compatíveis.

		Coeficientes	Erro Padrão	t Stat	valor-P	95% Inferiores	95% Superiores
Intersecção	β_0	4,71622	0,34900	13,51333	0,00088	3,60553	5,82690
Preço	β_1	-0,66216	0,08367	-7,91421	0,00421	-0,92843	-0,39589

Figura 23: Análise dos betas do exemplo 2 utilizando o Excel

Para agilizar a resolução das matrizes utilizamos o aplicativo **Matrix Operations**" (versão 2.08 da sssprog), gratuito, disponível para o sistema operacional **Android**.

4.3.2 *Exemplo 3 - Regressão Múltipla utilizando Notação Matricial*

Agora, vamos a um exemplo utilizando a notação matricial para regressão linear múltipla com duas variáveis independentes.

Exemplo 3 - Dado a tabela 20 temos os dados: Y = quantidade vendida de um produto, X_1 = o preço do produto e X_2 = o investimento com a divulgação do produto. Vamos determinar os coeficientes $\widehat{\beta}_0$, $\widehat{\beta}_1$ e $\widehat{\beta}_2$ e a equação de regressão linear múltipla.

Tabela 20: Quantidade em relação ao preço e investimento

Vendas	**Quantidade (kg)**	**Preço (R$)**	**Investimento (R$)**
1	55	100	550
2	70	90	630
3	90	80	720
4	100	70	700
5	90	70	625
6	105	70	735

Temos $k = 2$, pois nosso exemplo contém duas variáveis independentes (o preço e o investimento) e utilizando a equação 4.9 teremos

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = [X^t X]^{-1}[X^t Y]$$

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \left[\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 100 & 90 & 80 & 70 & 70 & 70 \\ 550 & 630 & 720 & 700 & 625 & 735 \end{bmatrix} \cdot \begin{bmatrix} 1 & 100 & 550 \\ 1 & 90 & 630 \\ 1 & 80 & 720 \\ 1 & 70 & 700 \\ 1 & 70 & 625 \\ 1 & 70 & 735 \end{bmatrix} \right]^{-1} \times$$

$$\times \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 100 & 90 & 80 & 70 & 70 & 70 \\ 550 & 630 & 720 & 700 & 625 & 735 \end{bmatrix} \cdot \begin{bmatrix} 55 \\ 70 \\ 90 \\ 100 \\ 90 \\ 105 \end{bmatrix}$$

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 6 & 480 & 3960 \\ 480 & 39200 & 313500 \\ 3960 & 313500 & 2638650 \end{bmatrix}^{-1} \times \begin{bmatrix} 510 \\ 39650 \\ 342575 \end{bmatrix}$$

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 171761/1830 & -697/1525 & -132/1525 \\ -697/1525 & 167/61000 & 11/30500 \\ -132/1525 & 11/30500 & 2/22875 \end{bmatrix}^{-1} \times \begin{bmatrix} 510 \\ 39650 \\ 342575 \end{bmatrix}$$

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 5699/61 \\ -303/305 \\ 197/1830 \end{bmatrix}$$

$$\begin{cases} \widehat{\beta}_0 = 93,42623 \\ \widehat{\beta}_1 = -0,99344 \\ \widehat{\beta}_2 = 0,107650 \end{cases}$$

Portanto, a equação de regressão linear múltipla é

$$\widehat{Y} = 93,42623 - 0,99344X_1 + 0,107650X_2$$

Utilizando o Excel para conferir os resultados (figura 24) verificamos que são totalmente compatíveis.

		Coeficientes	Erro Padrão	t Stat	valor-P	95% Inferiores	95% Superiores
Intersecção	β_0	93,426230	21,174263	4,412254	0,021602	26,040273	160,812186
Preço (R$)	β_1	-0,993443	0,114358	-8,687161	0,003210	-1,357379	-0,629506
Invest. (R$ mil)	β_2	0,107650	0,020436	5,267555	0,013335	0,042612	0,172688

Figura 24: Análise dos betas do exemplo 3 utilizando o Excel

4.4 ANÁLISE DE VARIÂNCIA (ANOVA)

4.4.1 *Soma de Quadrados*

Da mesma forma que utilizamos no MRLS teremos

$$SQT = \sum y^2 \tag{4.10}$$

$$SQE = \sum \widehat{y}^2 = \widehat{\beta}_1 \sum yx_1 + \widehat{\beta}_2 \sum yx_2 \tag{4.11}$$

$$SQR = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2 = SQT - SQE \tag{4.12}$$

4.4.2 *Graus de Liberdade*

Existe uma decomposição dos graus de liberdade associados (gl). A decomposição é a seguinte:

$$\begin{cases} \text{para SQE, gl} = k\text{, onde, } k \text{ é o número de variáveis explicativas,} \\ \text{para SQR, gl} = n \text{ - } (k + 1)\text{,} \\ \text{para SQT, gl} = n \text{ - 1.} \end{cases}$$

4.4.3 *Quadrado Médio*

O Quadrado Médio Explicado ou da Regressão (QME) é

$$QME = \frac{SQE}{gl} = \frac{SQE}{k} \tag{4.13}$$

O Quadrado Médio dos Resíduos (QMR) é

$$QMR = \sigma^2 = \frac{SQR}{gl} = \frac{SQR}{n - (k+1)} = \frac{SQT - SQE}{n - (k+1)} \tag{4.14}$$

4.4.4 *Tabela de Análise de Variância*

Podemos resumir a ANOVA nesta tabela que ajudará a analisar os dados.

Tabela 21: Tabela Resumo ANOVA

Fonte	**gl**	**Soma de Quadrados**	**Quadrado Médio**
Regressão (E)	k	$SQE = \sum \hat{y}^2 = \widehat{\beta_1} \sum yx_1 + \widehat{\beta_2} \sum yx_2$	$QME = \dfrac{SQE}{k}$
Resíduo (R)	$n-(k+1)$	$SQR = SQT - SQE$	$QMR = \dfrac{SQR}{n-(k+1)}$
Total (T)	$n-1$	$SQT = \sum y^2$	-

4.4.5 *Teste F*

Para determinar o teste de significância da regressão e verificar se há uma relação linear entre a variável resposta Y e algumas das variáveis regressora $X_1,\ X_2, \ldots,\ X_k$. Consideremos as hipóteses

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\ H_1 : \beta_j \neq 0 \quad \text{para qualquer } j \ = \ 1, \cdots, k \end{cases}$$

Se rejeitamos H_0, temos que ao menos uma variável explicativa $X_1,\ X_2, \ldots,\ X_k$ contribui significativamente para o modelo.

Sob H_0, temos pelo "Teorema - Distribuição de forma quadrática" que

$$\frac{SQE}{\sigma^2} \sim \chi^2_{(k)} \quad \text{e que} \quad \frac{SQR}{\sigma^2} \sim \chi^2_{(n-k-1)}.$$

Podemos utilizar a mesma equação (3.35) para determinar F_0.

$$F_0 = \frac{\dfrac{SQE}{p}}{\dfrac{SQR}{n-k-1}} = \frac{QME}{QMR} \ \sim \ F_{(k;\,n-k-1)}.$$

Portanto, rejeitamos H_0 se $F_0 > F_{(1-\alpha;k;\ n-k-1)}$ e se p-valor=P $[F_{k;\ n-k-1} > F_0] < \alpha$, em que α é o nível de significância considerado. Geralmente adotamos $\alpha = 5\% = 0,05$.

$$F_0 = \frac{QME}{QMR}$$

4.4.6 *Calculando ANOVA do Exemplo 1*

Agora vamos calcular os dados da ANOVA do nosso exemplo 1, conforme tabela 18.

Soma de Quadrados.

$$\begin{cases} SQT = \sum y^2 = 7.749,60 \quad \text{(valor constante na tabela)} \\ SQE = \widehat{\beta_1}\sum yx_1 + \widehat{\beta_2}\sum yx_2 = -0,2294 \cdot 16.820,8 + 5,4531 \cdot 1.923,6 = 6.630,89 \\ SQR = SQT - SQE = 7.749,60 - 6.630,89 = 1.118,71 \end{cases}$$

Graus de liberdade.

$$\begin{cases} gl \text{ (Regressão ou Explicativa)} = k = 2 \\ gl \text{ (Resíduo)} = n - (k+1) = 15 - (2+1) = 12 \\ gl \text{ (Total)} = n - 1 = 15 - 1 = 14 \end{cases}$$

Quadrado Médio.

$$\begin{cases} QME = \dfrac{SQE}{k} = \dfrac{6.630,89}{2} = 3.315,45 \\ \\ QMR = \dfrac{SQR}{n-(k+1)} = \dfrac{1.118,71}{12} = 93,23 \end{cases}$$

Teste F.

$$\begin{cases} F_0 = \dfrac{QME}{QMR} = \dfrac{3.315,45}{93,23} = 35,56 \end{cases}$$

Utilizando a tabela t (Figura 33) para $\alpha = 0,05 = 5\%$, obtemos que $F_{(0,95;2;12)} = (2,1788)^2 = 4,7472$

Logo,

$$F_0 = 35,56 > 4,7472 = F_{(0,95;2;12)}$$

Portanto, rejeitamos H_0 com um nível de confiança de 95%.

Utilizando o Excel para conferir os resultados (figura 25) verificamos que são totalmente compatíveis.

ANOVA

		gl	SQ	MQ	F	F de significação
Regressão	(E)	2	6630,73	3315,36	35,56	0,00091%
Resíduos	(R)	12	1118,87	93,24		
Total	(T)	14	7749,60			

Figura 25: Análise ANOVA do exemplo 1 utilizando o Excel

4.5 COEFICIENTE DE CORRELAÇÃO LINEAR E DE DETERMINAÇÃO

4.5.1 *Coeficiente de Determinação*

Podemos calcular o coeficiente de determinação (R^2) utilizando a equação 3.41

$$R^2 = \frac{SQE}{SQT} = 1 - \frac{SQR}{SQT},$$

portanto, podemos reescreve-la como

$$R^2 = \frac{\widehat{\beta_1} \sum yx_1 + \widehat{\beta_2} \sum yx_2}{\sum y^2} \tag{4.15}$$

4.5.2 *Coeficiente de Correlação Linear de Pearson*

Podemos calcular o coeficiente de correlação linear de Pearson (R)

$$R = \sqrt{R^2} = \sqrt{\frac{\widehat{\beta_1} \sum yx_1 + \widehat{\beta_2} \sum yx_2}{\sum y^2}} \tag{4.16}$$

4.5.3 *Coeficiente de Determinação Ajustado*

O valor do coeficiente de determinação ajustado $\bar{R}^2$ poderá ser calculado utilizando a equação (3.42), portanto teremos

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-(k+1)}\right) \cdot \left(1 - R^2\right) \tag{4.17}$$

4.5.4 *Calculando o Coeficientes do Exemplo 1*

Agora vamos calcular os coeficientes R, R^2 e $\bar{R}^2$ do nosso exemplo 1, conforme tabela 18.

Vamos recordar que temos $n = 15$ (observações) e $k = 2$ (variáveis independentes).

Utilizando a equação (4.15) para calcular o coeficiente de determinação (R^2)

$$R^2 = \frac{\widehat{\beta_1}\sum yx_1 + \widehat{\beta_2}\sum yx_2}{\sum y^2} = \frac{-0,2294 \cdot 16.820,8 + 5,4529 \cdot 1.923,6}{7.749,6}$$

$$R^2 = 0,8556$$

Utilizando a equação (4.16) para calcular o coeficiente de correlação linear (R)

$$R = \sqrt{R^2} = \sqrt{0,8556}$$

$$R = 0,9250$$

Utilizando a equação (4.17) para calcular o coeficiente de determinação ajustado ($\bar{R}^2$)

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-(k+1)}\right) \cdot (1 - R^2)$$

$$\bar{R}^2 = 1 - \left(\frac{15-1}{15-(2+1)}\right) \cdot (1 - 0,8556) = 1 - \frac{14}{12} \cdot 0,1444$$

$$\bar{R}^2 = 0,8315$$

Utilizando o Excel para conferir os resultados (figura 26) verificamos que são totalmente compatíveis.

Estatística de regressão	
R múltiplo	0,9250
R-Quadrado	0,8556
R-Quadrado Ajustado	0,8316
Erro Padrão	9,6561
Observações	15

Figura 26: Análise dos coeficientes do exemplo 1 utilizando o Excel

4.6 VARIÂNCIA

Como, em geral, σ^2 é desconhecido estimamos $Var(\widehat{\beta}_i)$ por $S_{\widehat{\beta}_i}$ que se obtém substituindo nas formulas anteriores σ^2 pelo seu estimador,

$$S^2 = \frac{SSE}{n-k-1}$$

Então,

$$S^2_{\widehat{\beta}_i} = \frac{SSE}{n-k-1}$$

4.7 EXERCÍCIO RESOLVIDO DE MRLM

1-) Uma empresa deseja contratar um vendedor e portanto pretende estudar a relação entre o volume de vendas (Y) efetuadas durante um dado período de tempo por um vendedor, os seus anos de experiência (X_1) e o seu score num teste de inteligência (X_2). Para isto ela utilizou os dados de dez vendedores conforme a tabela 25.

Tabela 22: Relação de vendas e anos de experiência e score no teste de inteligência dos vendedores

Vendedor	**Vendas (Y)**	**Anos de Experiência (X_1)**	**Score no teste (X_2)**
1	9	6	3
2	6	5	2
3	4	3	2
4	3	1	1
5	3	4	1
6	5	3	3
7	8	6	3
8	2	2	1
9	7	4	2
10	4	2	2

Determine:

a-) a tabela utilizando como modelo a Tabela 18.

b-) os estimadores dos parâmetros $\widehat{\beta_0}$, $\widehat{\beta_1}$ e $\widehat{\beta_2}$. Utilize as equações (4.6), (4.7) e (4.8) ou a equação (4.9) (forma matricial).

c-) a equação de regressão linear múltipla.

d-) o quadrado médio explicada ou da regressão (QME).

e-) o quadrado médio dos resíduos (QMR).

f-) o coeficiente de determinação (R^2).

g-) o coeficiente de correlação de Pearson (R).

h-) o coeficiente de determinação ajustado ($\bar{R}^2$).

Resolução:

a-) Vamos construir a tabela.

Tabela 23: Tabela da relação de vendas e anos de experiência e score no teste de inteligência dos vendedores

i	Y_i	X_1	X_2	y^2	x_1^2	x_2^2	yx_1	yx_2	x_1x_2
1	9	6	3	15,21	5,76	1,00	9,36	3,90	2,40
2	6	5	2	0,81	1,96	0,00	1,26	0,00	0,00
3	4	3	2	1,21	0,36	0,00	0,66	0,00	0,00
4	3	1	1	4,41	6,76	1,00	5,46	2,10	2,60
5	3	4	1	4,41	0,16	1,00	-0,84	2,10	-0,40
6	5	3	3	0,01	0,36	1,00	0,06	-0,10	-0,60
7	8	6	3	8,41	5,76	1,00	6,96	2,90	2,40
8	2	2	1	9,61	2,56	1,00	4,96	3,10	1,60
9	7	4	2	3,61	0,16	0,00	0,76	0,00	0,00
10	4	2	2	1,21	2,56	0,00	1,76	0,00	0,00
Soma	51	36	20	48,90	26,40	6,00	30,40	14,00	8,00
Média	5,10	3,60	2,00						

b-) Vamos determinar $\widehat{\beta}_1$ utilizando a equação (4.7)

$$\widehat{\beta}_1 = \frac{\Sigma yx_1 \Sigma x_2^2 - \Sigma yx_2 \Sigma x_1 x_2}{\Sigma x_1^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2}$$

$$\widehat{\beta}_1 = \frac{30,40 \cdot 6,00 - 14,00 \cdot 8,00}{26,40 \cdot 6,00 - (8,00)^2} = \frac{70,40}{94,40}$$

$$\widehat{\beta}_1 = 0,74576$$

Agora vamos determinar $\widehat{\beta}_2$ utilizando a equação (4.8)

$$\widehat{\beta}_2 = \frac{\Sigma yx_2 \Sigma x_1^2 - \Sigma yx_1 \Sigma x_1 x_2}{\Sigma x_1^2 \Sigma x_2^2 - (\Sigma x_1 x_2)^2}$$

$$\widehat{\beta}_1 = \frac{14,00 \cdot 26,40 - 30,40 \cdot 8,00}{26,40 \cdot 6,00 - (8,00)^2} = \frac{126,40}{94,40}$$

$$\widehat{\beta}_2 = 1,33898$$

Agora vamos determinar $\widehat{\beta}_0$ utilizando a equação (4.6)

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X}_1 - \widehat{\beta}_2 \bar{X}_2$$

$$\widehat{\beta}_0 = 5,10 - 0,74576 \cdot 3,60 - 1,33898 \cdot 2,00$$

$$\widehat{\beta}_0 = 5,10 - 2,684736 - 2,67796$$

$$\widehat{\beta}_0 = -0,2627$$

c-) A equação de regressão linear múltipla é

$$\widehat{Y} = -0,2627 - 0,74576X_1 + 1,33898X_2$$

d-) Para determinar o quadrado médio explicada ou da regressão (QME) utilizaremos a equação (3.33), mas antes devemos calcular a soma de quadrados explicada (SQE) utilizando a equação (3.29).

$$SQE = \widehat{\beta}_1 \sum yx_1 + \widehat{\beta}_2 \sum yx_2$$

$$SQE = 0,74576 \cdot 30,40 + 1,33898 \cdot 14,00$$

$$SQE = 22,671104 + 18,745720$$

$$SQE = 41,416824$$

$$QME = \frac{SQE}{k}$$

$$QME = \frac{41,416824}{2}$$

$$QME = 20,708412$$

e-) Para determinar o quadrado médio dos resíduos (QMR) utilizaremos a equação (3.34), mas antes devemos calcular a soma de quadrados total (SQT) utilizando a equação (3.28).

$$SQT = \sum y^2$$

$$SQT = 48,90$$

$$QMR = \frac{SQT - SQE}{n - (k+1)}$$
$$QMR = \frac{48,90 - 41,416824}{10 - (2+1)}$$
$$QMR = 1,0690$$

f-) Agora vamos calcular o coeficiente de determinação (R^2). Vamos recordar que temos $n = 10$ (observações) e $k = 2$ (variáveis independentes).

Utilizando a equação (4.15) para calcular o coeficiente de determinação (R^2)

$$R^2 = \frac{\widehat{\beta_1} \sum yx_1 + \widehat{\beta_2} \sum yx_2}{\sum y^2} = \frac{0,74576 \cdot 30,40 + 1,33898 \cdot 14,00}{48,90}$$

$$R^2 = 0,84697$$

g-) Utilizando a equação (4.16) para calcular o coeficiente de correlação linear (R)

$$R = \sqrt{R^2} = \sqrt{0,84697}$$

$$R = 0,9203$$

h-) o Utilizando a equação (4.17) para calcular o coeficiente de determinação ajustado ($\bar{R}^2$)

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-(k+1)}\right) \cdot \left(1 - R^2\right)$$

$$\bar{R}^2 = 1 - \left(\frac{10-1}{10-(2+1)}\right) \cdot (1 - 0,84697) = 1 - \frac{9}{7} \cdot 0,15303$$

$$\bar{R}^2 = 0,80325$$

4.8 EXERCÍCIOS PROPOSTOS DE MRLM

1-) Uma agroindústria deseja saber o custo de manutenção de sua frota de caminhões durante um determinado período. Utilizando um procedimento de amostragem foram coletados dados de quilometragem e a idade do caminhão. Na tabela 24 é possível visualizar esses valores.

Tabela 24: Relação do custo de manutenção pela quilometragem e idade do caminhão

Caminhão	**Custo de manutenção** (Y)	**Quilometragem** (X_1)	**Idade do caminhão** (X_2)
1	832	6.000	8
2	73	7.000	7
3	647	9.000	6
4	553	11.000	5
5	467	13.000	4
6	373	15.000	3
7	283	17.000	2
8	189	18.000	1
9	96	19.000	0

Determine:

a-) a tabela utilizando como modelo a Tabela (18).

b-) os estimadores dos parâmetros $\widehat{\beta}_0$, $\widehat{\beta}_1$ e $\widehat{\beta}_2$. Utilize as equações (4.6), (4.7) e (4.8) ou a equação (4.9) (forma matricial).

c-) a equação de regressão linear múltipla.

d-) o quadrado médio explicada ou da regressão (QME).

e-) o quadrado médio dos resíduos (QMR).

f-) o coeficiente de determinação (R^2).

g-) o coeficiente de correlação de Pearson (R).

h-) o coeficiente de determinação ajustado ($\bar{R}^2$).

2-) Um distribuidor de bebidas está analisando seu sistema de distribuição, sendo que seu interesse é atender o mais rápido possível um ponto de venda. O engenheiro industrial acredita que os dois fatores mais importantes são: número de caixas fornecidas e a distância do depósito ao posto de venda. Os dados coletados aparecem na tabela

Tabela 25: Relação do tempo de entrega com a quantidade de caixas e a distância

Tempo (h)	**Qtde de caixas** ($unid$)	**Distância** (km)
24	10	30
27	15	25
29	10	40
31	20	18
25	25	22
33	18	31
26	12	26
28	14	34
31	16	29
39	22	37
33	24	20
30	17	25
25	13	27
42	30	23
40	24	33

Determine:

a-) a tabela utilizando como modelo a Tabela (18).

b-) os estimadores dos parâmetros $\widehat{\beta}_0$, $\widehat{\beta}_1$ e $\widehat{\beta}_2$. Utilize as equações (4.6), (4.7) e (4.8) ou a equação (4.9) (forma matricial).

c-) a equação de regressão linear múltipla.

d-) o quadrado médio explicada ou da regressão (QME).

e-) o quadrado médio dos resíduos (QMR).

f-) o coeficiente de determinação (R^2).

g-) o coeficiente de correlação de Pearson (R).

h-) o coeficiente de determinação ajustado ($\bar{R}^2$).

5

ATIVIDADES COM OS ALUNOS

Neste capítulo teremos o desenvolvimento das atividades com os alunos.

Teremos inicialmente um plano de aula referente ao projeto proposto.

Será explicado a escolha do público alvo e o calendário de atividades e um pouco sobre a Unidade Escolar e por fim desenvolver uma série de atividades dirigidas com os alunos utilizando a dissertação como texto base.

5.1 PLANO DE AULA

5.1.1 *Período*

29/08 a 02/10/2016. Mínimo de 10 aulas, sendo 2 aulas semanais aplicados em 5 semanas.

5.1.2 *Dados de Identificação*

- Escola: ITB Prof. Antonio Arantes Filho
- Professor: Ricardo Will
- Professor colaborador: Leonardo R. Pfiszter
- Disciplinas: Matemática, Estatística, Física e Economia.
- Série: 3° anos do Ensino Médio.
- Turmas: Finanças (A e B) e Contabilidade (C).

- Período: Manhã

5.1.3 *Tema*

Introdução à Econometria no Ensino Médio: Aplicações da Regressão Linear.

5.1.4 *Objetivos*

Objetivos Gerais

- Permitir ao aluno desenvolver condições de entender as diversas aplicações e ser capaz de reconhecer o fenômeno linear e utilizar a regressão para fazer previsões.
- Desenvolver conhecimento e habilidades importantes para a realização de análise econométrica, permitindo confrontar as teorias com os dados sócio-econômicos.

Objetivos Específicos

- Preparar o aluno com o ferramental econométrico necessário para analisar dados sócio-econômicos e realizar análises de regressão.
- Desenvolver habilidades práticas de análise econométrica, através de ferramentas computacionais de análise estatística.

Competências e Habilidades

Competências [1]

- Interpretar informações de natureza científica e social obtidas da leitura de gráficos e tabelas, realizando previsão de tendência, extrapolação [2], interpolação [3] e interpretação.

1 Competências são um conjunto de habilidades harmonicamente desenvolvidas e que caracterizam por exemplo uma função/profissão específica: ser arquiteto, médico ou professor de química. As habilidades devem ser desenvolvidas na busca das competências.

2 Extrapolação é qualquer processo de obtenção dos valores de uma função fora de um intervalo, mediante o conhecimento de seu comportamento dentro desse intervalo.

3 Interpolação é o método que permite construir um novo conjunto de dados a partir de um conjunto discreto de dados pontuais previamente conhecidos.

Habilidades [4]

O que o aluno poderá aprender com este projeto.

- Relacionar o conteúdo a problemas do cotidiano e em situações interdisciplinares.
- Identificar os elementos que constituem uma equação.
- Reconhecer a incógnita e os símbolos matemáticos, numa dada situação, para representar diferentes problemas.
- Utilizar símbolos matemáticos para representar problemas do cotidiano em situações interdisciplinares.
- Associar uma tabela ao seu gráfico no plano cartesiano e construir um gráfico de dispersão.
- Identificar graficamente o coeficiente linear da função de 1º grau.
- Utilizar informações expressas em gráficos ou tabelas para fazer inferências.
- Analisar informações expressas em gráficos ou tabelas como recurso para a construção de argumentos.

5.1.5 *Conteúdos*

- Noções de Estatística.
- Conceitos de Econometria: introdução, objetivos, análise econométrica e metodologia econométrica.
- Modelo de Regressão Linear Simples: introdução, diagrama de dispersão, estimação dos parâmetros pelo Método dos Mínimos Quadrados Ordinários (MQO), estimação dos resíduos, análise de variância, coeficiente de correlação e inferência estatística.
- Modelo de Regressão Linear Múltipla: introdução, estimação dos parâmetros pelo MQO e pela notação matricial, análise de variância, coeficiente de correlação.

4 As habilidades estão associadas ao saber fazer: ação física ou mental que indica a capacidade adquirida. Assim, identificar variáveis, compreender fenômenos, relacionar informações, analisar situações-problema, sintetizar, julgar, correlacionar e manipular são exemplos de habilidades.

5.1.6 *Metodologia*

O conteúdo será abordado por meio de aulas expositivas e práticas em sala de aula e laboratório de informática, utilizaremos videoaulas e postagens de materiais em site do professor. Trabalhos prático individuais e em duplas e listas de exercícios resolvidos que complementam e reforçam o conteúdo. O programa econométrico utilizado nas aulas práticas e recomendado para o trabalho será o Excel, já instalado na Unidade Escolar, no entanto não dispensa os cálculos. Outros materiais, como base de dados e listas de exercícios, foram disponibilizados no site do professor: https://sites.google.com/site/regressaolinear/home

5.1.7 *Recursos Disponíveis*

- Lousa e giz.
- Laboratório de informática (durante o período de aula).
- Notebooks particulares (fora do período de aula).
- Internet.
- Aplicativos de smartphone.

5.1.8 *Avaliação*

Será apenas de orientação e verificação do processo de ensino aprendizagem e não valendo para composição de nota nas disciplinas, no entanto, os professores irão utilizar como bonificação na média final.

5.1.9 *Referências*

As aulas terão como texto base a dissertação. Ela foi disponibilizada por capítulos e postada no site. A única parte retirada foi a resolução das questões propostas. Também foi postada toda a bibliografia da dissertação.

5.2 O PÚBLICO ALVO

Foram selecionados os alunos dos 3° anos dos cursos de: Finanças e Contabilidade que são ministrados no ITB Prof. Antonio Arantes Filho, localizado na Estrada das Pitas, 799 - Parque Viana, em Barueri-SP.

A escolha destes alunos são em virtude do curso de Finanças ter em sua grade curricular a disciplina de Estatística e Economia. A turma de Contabilidade (ver figura 27) foi voluntária apesar de não ter aulas de Economia.

Também pude contar com a colaboração do Prof. Leonardo R. Pfiszter que ministra no curso de Finanças aulas de Economia, onde cedeu espaço, tempo e conhecimento.

Figura 27: Alunos do ITB Prof. Antonio Arantes Filho resolvendo as atividades em sala de aula.

5.3 A UNIDADE ESCOLAR

Inicialmente solicitei autorização da Gestão Escolar para poder realizar as atividades como os alunos e utilizar o nome da Unidade Escolar.

O **ITB** (Instituto Técnico de Barueri) é subordinado a FIEB - Fundação Instituto de Educação de Barueri, que é mantida pela Prefeitura Municipal de Barueri. São seis unidades do ITB distribuídas pela cidade com diversos cursos técnicos conforme relação a seguir:

- ITB Brasílio Flores de Azevedo – Jd. Belval
  Cursos: Edificações, Eletroeletrônica, Informática e Telecomunicações
- ITB Prof. Munir José – Jd. Paulista
  Cursos: Administração, Redes de Computadores, Recursos Humanos, Secretariado e Segurança do Trabalho
- ITB Profª. Maria Sylvia Chaluppe Mello – Engenho Novo
  Cursos: Análises Clínicas, Química, Enfermagem, Farmácia e Informática para Internet
- ITB Prof. Hercules Alves de Oliveira – Jd. Mutinga
  Cursos: Design de Interiores e Publicidade
- ITB Prof. Moacyr Domingos Sávio Veronezi – Pq. Imperial
  Cursos: Logística, Manutenção e Suporte em Informática e Segurança do Trabalho
- ITB Prof. Antonio Arantes Filho – Pq. Viana
  Cursos: Contabilidade, Finanças, Serviços Públicos e Hospedagem.

A unidade mais antiga é o ITB Brasílio Flores de Azevedo que completou 22 anos e a unidade do ITB Prof. Antonio Arantes Filho (figura 28) é uma das unidades mais nova com seis anos.

Figura 28: ITB Prof. Antonio Arantes Filho

Os cursos são oferecidos concomitante ao Ensino Médio com uma carga horária anual de 1280 horas, distribuídas em 32 horas semanais e distribuídas entre 16 a 18 disciplinas por ano. O ano letivo é dividido em três trimestres.

Devido à grande quantidade de disciplinas na grade escolar a carga horária de matemática fica reduzida a 3 aulas semanais de 50 minutos cada.

No ITB Prof. Antonio Arantes Filho as aulas são ministradas das 07:00 horas às 13:10 horas. O prédio é compartilhado com a unidade do EMEF Prof. Renato Rosa (Escola Municipal do Ensino Fundamental) que não faz parte da FIEB e é subordinada a Secretaria de Educação de Barueri. Os alunos concluintes do Ensino Fundamental das escolas municipais tem prioridade ao acesso aos ITBs e as vagas remanescentes são oferecidas através de vestibulinho

5.4 CALENDÁRIO E CRONOGRAMA DAS ATIVIDADES

Devido as avaliações trimestrais obrigatórias da escola que ocorreram na segunda quinzena de agosto ficou acertado com a gestão escolar que as atividades somente poderiam iniciar ao término destas avaliações, portanto, o nosso calendário das atividades ficou do dia 29/08 a 02/10/2016.

As atividades foram divididas em cinco semanas, sendo duas por semana, tendo um total de dez atividades.

Os alunos foram informados do projeto e convidados a participar. Como estas atividades seriam extras e sem prejuízo as aulas diárias, portanto as aulas ocorreram uma dentro do período letivo e outra ao término do período. E como haviam alunos interessados e que não poderiam ficar devido a trabalho, estágio e ou outros cursos, a solução foi criar um site onde foram postados todas as atividades e as orientações e inclusive vídeos aulas disponíveis no YouTube.

Os alunos puderam entregar os trabalhos pessoalmente ou digitalizados e enviados por email.

O cronograma das atividades solicitadas foram:

- **Semana 1 - 29/08 a 04/09/16**
  Atividade 1 - Responda o Questionário de Pesquisa antes de iniciar os estudos. Retire o formulário com o professor ou imprima diretamente do site.
  Atividade 2 - Leiam o capítulo sobre Noções de Estatística e refaçam os exercícios, também tem vídeo aulas, não será necessário entregar.
- **Semana 2 - 05/09 a 11/09/16**
  Atividade 3 - Leiam o capítulo de Conceitos de Econometria e assistam os vídeos

indicados em Diagrama de Dispersão e Keynes e Propensão Marginal ao Consumo.
Atividade 4 - Resolva as questões indicadas (no capítulo de Conceitos de Econometria) e entregue ao professor.

- **Semana 3 - 12/09 a 18/09/16**
  Atividade 5 - Leiam o capítulo de Modelo de Regressão Linear Simples (MRLS) e assistam os vídeos indicados.
  Atividade 6 - Refaçam o exercício 1 proposto de MRLS. Entregue o exercício refeito ao professor com as observações de cada item (por exemplo facilidade, dificuldade, utilizou calculadora ou outros recursos, consultou colegas ou outras pessoas, teve que pesquisar mais, onde? E outras que poderá otimizar o processo de aprendizagem). Esta atividade poderá ser em dupla, no entanto as atividades 7 e 8 deverão ser individuais.

- **Semana 4 - 19/09 a 25/09/16**
  Atividade 7 - Resolva o exercício 02 de MRLS e entregue ao professor.
  Atividade 8 - Resolva o exercício 03 de MRLS e entregue ao professor.

- **Semana 5 - 26/09 a 02/10/16**
  Atividade 9 - Leiam o capítulo de Modelo de Regressão Linear Múltipla (MRLM) e assistam os vídeos indicados.
  Atividade 10 - Resolva o exercício 01 ou 02 ou 03 das de MRLM e entregue ao professor. Também faça um texto comentando sobre as dificuldades que teve (positivas e negativas) sobre a assunto abordado nestas cinco semanas. Tente fazer uma auto avaliação. Dê também sugestões que possam ajudar no processo de ensino aprendizagem.

Todas as atividades, textos, videoaulas e orientações foram disponibilizadas no site: https://sites.google.com/site/regressaolinear/home (figura 29)

Figura 29: Tela inicial site trabalhado com os alunos

5.5 AS ATIVIDADE DE 1 E 2.

A Atividade 1, que é o questionário (figura 32 - Apêndice A.2), tem a intenção de verificar os conhecimentos prévios dos alunos nas atividades que se seguirão. Para isto foi elaborada 10 questões simples, onde apenas uma é aberta. Tendo por finalidade dar uma noção do rumo inicial do processo e contribuir com maior eficiência nas aulas, sendo que estas são muito limitadas.

Responderam o questionário um total de 57 alunos, sendo 36 do curso de Finanças e 21 do curso de Contabilidade.

1-) Você estudou (ou está estudando) Estatística no Ensino Médio? Caso afirmativo: por quanto tempo?

Sim	Não
41 (72%)	16 (28%)

Todos que responderam sim informaram que estudaram Estatística por 2 semestres. Sendo que foi no primeiro ano do ensino médio.

2-) Você sabe o que é Teste de Hipóteses em Estatística?

Sim	Não
12 (21%)	45 (79%)

3-) Você sabe o que é Intervalo de Confiança em Estatística?

Sim	Não
13 (23%)	44 (77%)

4-) Você sabe interpretar gráficos lineares (por exemplo: vendas x tempo)?

Sim	Não
54 (95%)	03 (5%)

5-) Você sabe construir um gráfico linear utilizando uma tabela de dados (por exemplo: vendas x tempo)?

Sim	Não
45 (79%)	12 (21%)

6-) Você já construiu um gráfico de dispersão?

Sim	Não
04 (7%)	53 (93%)

Caso negativo. - Você sabe o que é um gráfico de dispersão?

Sim	Não
10 (18%)	47 (82%)

7-) Dado uma função do 1º grau você sabe interpretar quem é o coeficiente linear (intercepto) e o coeficiente angular?

Sim	Não
23 (40%)	34 (60%)

8-) Se for dado um gráfico do 1º grau você sabe determinar a sua função do 1º grau (por exemplo: dado um gráfico da posição x tempo, em movimento uniforme, isto é, velocidade constante)?

Sim	Não
25 (44%)	32 (56%)

9-) Você sabe o que é Propensão Marginal a Consumir e a Poupar?

Sim	Não
13 (23%)	44 (77%)

10-) Qual seu conhecimento hoje sobre Regressão Linear Simples?

Muito superficial	Absolutamente nada
06 (11%)	51 (89%)

Podemos verificar que a maior parte dos alunos já estudaram Estatística, no entanto, poucos sabem o que é um teste de hipóteses e intervalo de confiança. O que demonstra que os alunos sabem apenas os conceitos básicos de estatística.

Quase a totalidade dos alunos sabem interpretar gráficos lineares, mas apenas metade deles sabem construir os gráficos. E a maior parte dos alunos nunca construíram e nem se quer sabem o que é um gráfico de dispersão.

Praticamente a metade dos alunos não sabem determinar os coeficientes: linear e angular de um gráfico linear, sendo abordado no primeiro ano do ensino médio em Matemática e em Física.

Como já era de se esperar a maior parte dos alunos não sabem o que propensão marginal ao consumo e a poupar e principalmente o que é Regressão Linear Simples.

Na primeira semana foi a entrega do questionário e a leitura do capítulo de Noções de Estatística, para isto, tivemos duas aulas de estatística e refizemos os exercícios indicados no capítulo.

No decorrer das aulas percebi que a maioria dos alunos não tinham lido e refeito as questões. Fato que tive que retomar as explicações e reiterar a importância do capítulo de Noções de Estatística.

5.6 AS ATIVIDADE DE 3 E 4.

Usando a análise dos questionários na segunda semana foi dado ênfase na construção de tabelas, gráficos de dispersão, interpretação gráfica e a determinação dos coeficientes: angular e linear.

Os alunos passaram a perguntar mais e tirar dúvidas, dando para perceber que já estavam mais por dentro do assunto, pois as perguntas eram mais específicas.

O tempo para esta atividade deve ser revisto e estendido, pois as dúvidas geraram atraso na entrega das resoluções dos exercícios e ocasionando um efeito de bola de neve, sendo que alguns alunos acreditaram que poderiam resolver as atividades rapidamente, na véspera ao dia da entrega das atividades.

Nesta atividade foi recomendado que utilizassem o Excel para determinar a função keynesiana, construir a tabela e o gráfico, para isto utilizamos o laboratório de informática (figuras 30 e 31), no entanto, solicitei que fizessem os cálculos de determinação dos parâmetros para confrontar com os valores que o programa forneceu. O aproveitamento nesta atividade foi espetacular, inclusive com perguntas via email, sendo que alguns alunos enviam tarde da noite e cobravam a resposta na manhã do dia seguinte.

Figura 30: Alunos do ITB Prof. Antonio Arantes Filho resolvendo as atividades na sala de informática.

Figura 31: Alunos do ITB Prof. Antonio Arantes Filho resolvendo as atividades na sala de informática.

5.7 AS ATIVIDADE DE 5 E 6

A atividade 5 era apenas a leitura do capítulo de MRLS e assistir aos vídeos indicados, caso fosse necessário.

Na atividade 6 como as atividades atrasaram permiti que fizessem em dupla e assim a maioria se ajustou ao prazo. Os trabalhos passaram a compor nota de participação na disciplina de Economia do Prof. Leonardo e com isto os alunos se animaram. Esta atividade tinha como objetivo preparar os alunos para as atividades 7 e 8, portanto eles deveriam apenas refazer o exercício (número 01) já resolvido e tirar as dúvidas.

Aconselhei que fizessem um tabela bem organizada, com as fórmulas corretas e salvassem em um arquivo, pois serviria de esqueleto para as demais atividades.

Também solicitei que fizessem observações sobre a atividade como: facilidade, dificuldade, utilizou calculadora ou outros recursos, consultou colegas ou outras pessoas, teve que pesquisar mais, onde? E outras que poderá otimizar o processo de aprendizagem.

A maioria dos alunos se preocuparam em apenas responder as questões de cálculos e acabaram não respondendo as observações. Os que responderam, em geral, informaram que o nível de dificuldade foi médio, sendo que a maior dificuldade foi a inferência estatística, onde precisaram de ajuda. O uso do Excel e fazer em dupla facilitou muito e também tirar dúvidas com os professores.

5.8 AS ATIVIDADE DE 7 E 8

Estas duas atividades eram semelhantes as anteriores, referente a MRLS, e deveriam ser feitas individualmente.

Os alunos tiveram mais facilidade em resolver estas atividades em virtude utilizarem a dica de utilizar a tabela do Excel do exercício anterior como um esqueleto para facilitar as resoluções dos exercícios.

Também a terminologia utilizada passou a ser mais conhecida e compreendida pelos alunos.

A resolução das atividades 7 e 8 feita por um aluno de Contabilidade e enviada por email está no Anexo B.

5.9 AS ATIVIDADE DE 9 E 10

A atividade 9 foi foi a leitura do capítulo de MRLM e assistir aos vídeos indicados, caso fosse necessário.

Na atividade 10 o aluno deveria escolher apenas um exercício para resolver entre os três propostos de MRLM. Também solicitei um texto comentando sobre as dificuldades que teve (positivas e negativas) sobre a assunto abordado nestas cinco semanas e uma auto avaliação. E sugestões que possam ajudar no processo de ensino aprendizagem.

A pergunta que mais foi feita foi sobre o número de variáveis explicativas.

Apenas um aluno resolveu os parâmetros pela notação matricial, sendo que os demais alunos fizeram pelo Método dos Mínimos Quadrados Ordinários em virtude de ter fórmulas e ter aprendido fazer a tabela.

Um aluno fez o seguinte comentário: "Que a primeira atividade foi fácil (ele está se referindo a atividade 4), pois tinha vídeo-aulas que explicava tudo o que tinha que fazer. A segunda (ele se refere a atividade 6, 7 e 8) foi difícil e muto longa e levou muito tempo para entender. A última foi mais simples, pois foi mais curta e muito semelhante as anteriores."

6

CONCLUSÃO

O trabalho proposto com os alunos do 3° ano do ensino médio do ITB Prof. Antonio Arantes Filho foi desenvolvido conforme o plano de aula obedecendo o calendário e a sequência proposta das atividades.

A gestão escolar colaborou em todos os aspectos, desde a autorização, utilização espaço físico e dos equipamentos, inclusive questionando os alunos sobre o projeto e incentivando-os. Também pude contar com o apoio do professor Leonardo, de Economia, inicialmente com dicas sobre o tema, de livros, apostilas, programas de análise de dados (Stata, Excel, Gretl e Action) e posteriormente com uso de nota de participação.

O período escolhido para a realização das atividades não ajudou muito, pois os alunos tinham terminado um período de provas obrigatórias e estavam cansados e acabou contribuindo na demora de uma dedicação ideal, portanto, deve ser levado em conta todas as influências internas e externas para a escolha do período do projeto e uma maior integração dos professores.

Nas escolas de Barueri os professores tem 1/3 da carga horária fora da sala de aula, o que facilitou muito, pois tinha um horário reservado na biblioteca para aulas extras, sendo uma por semana, mas poucos alunos podiam ficar até mais tarde na escola por motivos de trabalho, cursos, estágios ou outros.

Como a unidade escolar possui laboratório de informática com computadores (um por aluno) e um técnico de informática e a maior parte dos alunos conhecer o básico do Excel tive mais facilidade nas aulas e na realização das atividades como os alunos, como por exemplo: construção de tabelas e gráficos.

O tempo entre as atividades deve ser maior, pois a maioria dos alunos tiveram dificuldades em entregar no prazo, pois vários fatores interferiram: cansaço pós prova, tempo para entendimento, trabalho de outros professores e outros.

Perguntado a eles sobre a interpretação dos dados percebi que haviam adquirido habilidades de entender, interpretar, reconhecer o fenômeno linear e realizar previsões econométricas. Inclusive um aluno até fez um paralelo da cultura dos americanos com os brasileiros sobre poupar dinheiro pensando no futuro dos filhos e da previdência privada citando os postulados de Keynes sobre propensão marginal ao consumo e a poupar e que levaria como lição para sua vida.

Na última atividade sobre o Modelo de Regressão Linear Múltipla reduzi a quantidade de itens nas perguntas, pois os alunos já haviam atingido os objetivos propostos.

Este projeto sendo incluído no planejamento anual e feito com apoio dos demais professores irá contribuir muito no ensino aprendizagem dos alunos e terá um resultado extremamente satisfatório, inclusive sendo possível incluir alguma saída de atividade de campo relacionada ao assunto.

Para este projeto as aulas de Matemática Discreta, Fundamentos de Cálculo, Geometria Analítica e Recursos Computacionais do PROFMAT foram de extrema importância, pois em todos os momentos de pesquisa e elaboração da dissertação utilizei os aprendizados e repassei os meus alunos.

A

APÊNDICE A

A.1 EFEITO MULTIPLICADOR DE RENDA

Tabela 26: Efeito multiplicador de renda com PMC = 0,7169

Agente	**Recebimento**	**Circulante**	**Poupança**
turista	-	100,00	-
hospedagem no hotel	100,00	71,69	28,31
fornec. de carnes	71,69	51,39	20,30
criador de animais	51,39	36,84	15,55
veterinário	36,84	26,41	10,43
recepcionista	26,41	18,93	7,48
mercearia	18,93	13,57	5,36
entregador	13,57	9,72	3,85
lanchonete	9,72	6,97	2,75
fornec. mat. limpeza	6,97	5,00	1,97
fornec. prod. químicos	5,00	3,58	1,42
combustível	3,58	2,56	1,02
frentista	2,56	1,83	0,73
supermercado	1,83	1,31	0,52
fornec. verduras	1,31	0,94	0,37
fertilizantes	0,94	0,67	0,27
embalagens	0,67	0,48	0,19
frete	0,48	0,35	0,13
auto peças	0,35	0,25	0,10
motoboy	0,25	0,18	0,07
padaria	0,18	0,13	0,05
fornec. farinha	0,13	0,09	0,04
motorista	0,09	0,07	0,02
feirante	0,07	0,04	0,03
mesada filho	0,04	0	0,03
-	**Totais**	**353,00**	**100,00**

A.2 QUESTIONÁRIO DE PESQUISA SOBRE CONHECIMENTOS PRÉVIOS

Questionário de Pesquisa sobre Conhecimentos Iniciais

Nome: ____

Curso que está frequentando no ITB Prof. Antonio Arantes Filho: ______________________________ 3º ano

Idade: ___________ data: _____ / _____ / 2016

Este questionário tem a finalidade de conhecer o perfil dos alunos sobre seus conhecimentos iniciais sobre os tópicos que iremos trabalhar nas atividades que iremos desenvolver no Trabalho sobre Introdução à Econometria no Ensino Médio: Aplicações da Regressão Linear.

São apenas 10 questões. Marque um **X** nos quadros (abaixo das afirmações).

1) Você estudou (ou está estudando) Estatística no Ensino Médio? Caso afirmativo: por quanto tempo?

Sim	**Não**

1 semestre	**2 semestres**	**3 semestres**	**4 semestres**	**Mais de 4 semestres**

2) Você sabe o que é Teste de Hipóteses em Estatística?

Sim	**Não**

3) Você sabe o que é Intervalo de Confiança em Estatística?

Sim	**Não**

4) Você sabe interpretar gráficos lineares (por exemplo: vendas x tempo)?

Sim	**Não**

5) Você sabe construir um gráfico linear utilizando uma tabela de dados (por exemplo: vendas x tempo)?

Sim	**Não**

6) Você já construiu um gráfico de dispersão? Caso negativo: Você sabe o que é um gráfico de dispersão?

Sim	**Não**

Sim	**Não**

7) Dado uma função do 1º grau você sabe interpretar quem é o coeficiente linear (intercepto) e o coeficiente angular?

Sim	**Não**

8) Se for dado um gráfico do 1º grau você sabe determinar a sua função do 1º grau (por exemplo: dado um gráfico da posição x tempo, em movimento uniforme, isto é, velocidade constante)?

Sim	**Não**

9) Você sabe o que é Propensão Marginal a Consumir e a Poupar?

Sim	**Não**

10) Qual seu conhecimento hoje sobre Regressão Linear Simples?

Acesse o site: https://sites.google.com/site/regressaolinear/home

Figura 32: Questionário de Pesquisa sobre Conhecimentos Prévios

ANEXO A

1 TABELA DE DISTRIBUIÇÃO T - STUDENT

gl	Teste Unilateral								
	15%	10%	5%	2,5%	2%	1%	0,5%	0,1%	0,05%
	Teste Bilateral								
	30%	20%	10%	5%	4%	2%	1%	0,2%	0,1%
1	1,9626	3,0777	6,3137	12,7062	15,8945	31,8210	63,6559	318,2888	636,5776
2	1,3862	1,8856	2,9200	4,3027	4,8487	6,9645	9,9250	22,3285	31,5998
3	1,2498	1,6377	2,3534	3,1824	3,4819	4,5407	5,8408	10,2143	12,9244
4	1,1896	1,5332	2,1318	2,7765	2,9985	3,7469	4,6041	7,1729	8,6101
5	1,1558	1,4759	2,0150	2,5706	2,7565	3,3649	4,0321	5,8935	6,8685
6	1,1342	1,4398	1,9432	2,4469	2,6122	3,1427	3,7074	5,2075	5,9587
7	1,1192	1,4149	1,8946	2,3646	2,5168	2,9979	3,4995	4,7853	5,4081
8	1,1081	1,3968	1,8595	2,3060	2,4490	2,8965	3,3554	4,5008	5,0414
9	1,0997	1,3830	1,8331	2,2622	2,3984	2,8214	3,2498	4,2969	4,7809
10	1,0931	1,3722	1,8125	2,2281	2,3593	2,7638	3,1693	4,1437	4,5868
11	1,0877	1,3634	1,7959	2,2010	2,3281	2,7181	3,1058	4,0248	4,4369
12	1,0832	1,3562	1,7823	2,1788	2,3027	2,6810	3,0545	3,9296	4,3178
13	1,0795	1,3502	1,7709	2,1604	2,2816	2,6503	3,0123	3,8520	4,2209
14	1,0763	1,3450	1,7613	2,1448	2,2638	2,6245	2,9768	3,7874	4,1403
15	1,0735	1,3406	1,7531	2,1315	2,2485	2,6025	2,9467	3,7329	4,0728
16	1,0711	1,3368	1,7459	2,1199	2,2354	2,5835	2,9208	3,6861	4,0149
17	1,0690	1,3334	1,7396	2,1098	2,2238	2,5669	2,8982	3,6458	3,9651
18	1,0672	1,3304	1,7341	2,1009	2,2137	2,5524	2,8784	3,6105	3,9217
19	1,0655	1,3277	1,7291	2,0930	2,2047	2,5395	2,8609	3,5793	3,8833
20	1,0640	1,3253	1,7247	2,0860	2,1967	2,5280	2,8453	3,5518	3,8496
21	1,0627	1,3232	1,7207	2,0796	2,1894	2,5176	2,8314	3,5271	3,8193
22	1,0614	1,3212	1,7171	2,0739	2,1829	2,5083	2,8188	3,5050	3,7922
23	1,0603	1,3195	1,7139	2,0687	2,1770	2,4999	2,8073	3,4850	3,7676
24	1,0593	1,3178	1,7109	2,0639	2,1715	2,4922	2,7970	3,4668	3,7454
25	1,0584	1,3163	1,7081	2,0595	2,1666	2,4851	2,7874	3,4502	3,7251
26	1,0575	1,3150	1,7056	2,0555	2,1620	2,4786	2,7787	3,4350	3,7067
27	1,0567	1,3137	1,7033	2,0518	2,1578	2,4727	2,7707	3,4210	3,6895
28	1,0560	1,3125	1,7011	2,0484	2,1539	2,4671	2,7633	3,4082	3,6739
29	1,0553	1,3114	1,6991	2,0452	2,1503	2,4620	2,7564	3,3963	3,6595
30	1,0547	1,3104	1,6973	2,0423	2,1470	2,4573	2,7500	3,3852	3,6460
35	1,0520	1,3062	1,6896	2,0301	2,1332	2,4377	2,7238	3,3400	3,5911
40	1,0500	1,3031	1,6839	2,0211	2,1229	2,4233	2,7045	3,3069	3,5510
50	1,0473	1,2987	1,6759	2,0086	2,1087	2,4033	2,6778	3,2614	3,4960
60	1,0455	1,2958	1,6706	2,0003	2,0994	2,3901	2,6603	3,2317	3,4602
120	1,0409	1,2886	1,6576	1,9799	2,0763	2,3578	2,6174	3,1595	3,3734
+∞	1,0364	1,2816	1,6449	1,9600	2,0537	2,3264	2,5758	3,0902	3,2905

Figura 33: Tabela de Distribuição t - Student [4]

2 TABELA DE DISTRIBUIÇÃO NORMAL ESCORES NEGATIVOS

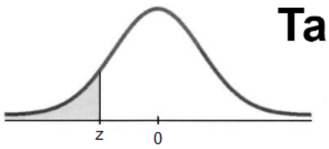


Tabela da Distribuição Normal

Área Acumulada à Esquerda

(Escores de z negativos)

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,6	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Figura 34: Tabela de Distribuição Normal Escores Negativos

3 TABELA DE DISTRIBUIÇÃO NORMAL ESCORES POSITIVOS

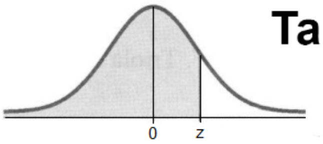


Tabela da Distribuição Normal

Área Acumulada à Direita

(Escores de z positivos)

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Figura 35: Tabela de Distribuição Normal Escores Positivos

ANEXO B

1 ATIVIDADE 7 RESOLVIDA PELO ALUNO

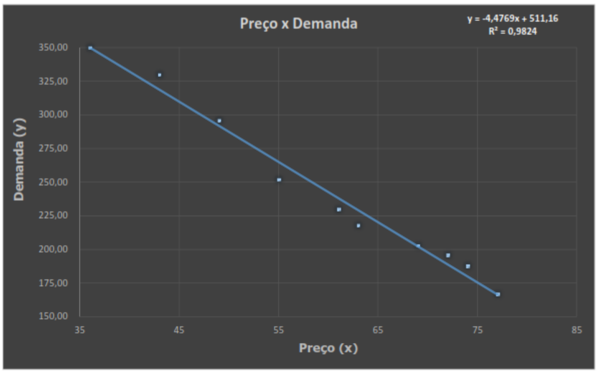
fieb	ITB – PROF. ANTÔNIO ARANTES FILHO			ITB Instituto Técnico de Barueri Prof. Antonio Arantes Filho
	Aluno: Jenilson Melo Nascimento		**Número:** 14	
	3ª série:	**Curso:** Contabilidade	**Turma:** C	
Professor: Ricardo Will		**Disciplina:** Física	**Período:** Manhã	
		3 º TRIMESTRE	**Data:** 20 / set / 2016	**Nota**

Atividade 7 – Modelo de Regressão Linear Simples

02. Uma empresa resolveu estudar a variação da demanda (em unidades) de seu produto em função do preço de venda (em reais) praticado. Para isso, foram coletados os seguintes dados:

Tabela 16: Variação da demanda do produto em função do preço de venda

Preço (X)	Demanda (Y)
36	350,00
43	330,00
49	296,00
55	252,00
61	230,00
63	218,00
69	203,00
72	196,00
74	188,00
77	167,00



Determine:

a) Os estimadores Sxx, Sxy, Syy

	Preço (X)	Demanda (Y)	(Xi - X)²	(Yi - Y)²	(Xi - X)(Yi - Y)
	36	350,00	571,21	11.449,00	- 2.557,30
	43	330,00	285,61	7.569,00	- 1.470,30
	49	296,00	118,81	2.809,00	- 577,70
	55	252,00	24,01	81,00	- 44,10
	61	230,00	1,21	169,00	- 14,30
	63	218,00	9,61	625,00	- 77,50
	69	203,00	82,81	1.600,00	- 364,00
	72	196,00	146,41	2.209,00	- 568,70
	74	188,00	198,81	3.025,00	- 775,50
	77	167,00	292,41	5.776,00	- 1.299,60
Totais	599,00	2.430,00	1.730,90	35.312,00	- 7.749,00
Médias	59,90	243,00			

$Sxx = \sum(X_i - X)^2$
$Sxx = 1.730,90$

$Syy = \sum(Y_i - Y)^2$
$Syy = 35.312,00$

$Sxy = \sum(X_i - X)(Y_i - Y)$
$Sxy = - 7.749,00$

b) Os parâmetros β_o e β_1.

$$\beta_1 = \frac{Sxy}{Sxx} = \frac{-7.749,00}{1.730,90} = -4,4769$$

$\beta_0 = Y - \beta_1 X$
$\beta_0 = 243 - (-4,4769)(59,90)$
$\beta_0 = 511,16$

Figura 36: Atividade 7 resolvida pelo aluno e enviada via email - parte 1

c) A equação de regressão linear simples

$Y = -4,4769x + 511,16$

d) A soma dos quadrados dos resíduos (SQR)

$\mathbf{SQR = Syy - \beta_1 * Sxy}$

$SQR = 35.312 - (-4,4769 * -7.749)$

$SQR = 620,50$

e) O quadrado médio dos resíduos (QMR)

$\mathbf{QMR} = 6^2$

$\mathbf{QMR} = \frac{SQR}{n-2}$

$QMR = \frac{620,50}{8}$

$QMR = 77,56$

f) O desvio padrão dos resíduos

$Ϭ = \sqrt{6^2}$

$Ϭ = \sqrt{77,56}$

$Ϭ = 8,8068$

g) A variância de β_0 e β_1

$\mathbf{var\ (\beta_0)} = \sqrt{\frac{6^2}{\beta 0}}$

$\mathbf{var\ (\beta_0) = Ϭ^2 * \frac{1}{n} + \frac{X^2}{Sxx}}$

$var\ (\beta_0) = 77,56 * (\frac{1}{10} + \frac{(599)^2}{1730,90})$

$var\ (\beta_0) = 16.085,29$

$\mathbf{var\ (\beta_1)} = \sqrt{\frac{6^2}{\beta 1}}$

$\mathbf{var\ (\beta_1)} = \frac{6^2}{Sxx}$

$var\ (\beta_1) = \frac{77,56}{1730,90}$

$var\ (\beta_1) = 0,04481$

h) O desvio padrão de β_0 e β_1

$Ϭ_{\beta 0} = \sqrt{\frac{6^2}{\beta 0}}$

$Ϭ_{\beta 0} = \sqrt{16.085,29}$

$Ϭ_{\beta 0} = 126,83$

$Ϭ_{\beta 1} = \sqrt{\frac{6^2}{\beta 1}}$

$Ϭ_{\beta 1} = \sqrt{0,04481}$

$Ϭ_{\beta 1} = 0,21168$

Figura 37: Atividade 7 resolvida pelo aluno e enviada via email - parte 2

i) A análise da variância (ANOVA) completando a tabela

ANOVA

	gl	*SQ*	*MQ*	*F*	*F de significação*
Regressão	1	34691,20169	34691,20169	447,052783	2,63056E-08
Resíduo	8	620,798313	77,59978913		
Total	9	35312			

SQT = Syy
SQT = 35.312

SQE = QME
SQE = SQT – SQR
SQE = 35.312 – 620,50
SQE = 34.691,50

j) O teste F para um nível de confiança de 95%

$$F_0 = \frac{QME}{QMR} \qquad F_0 = \frac{34.691,5}{77,56}$$

$$F_0 = 447,29$$

k) O coeficiente de correlação Pearson.

$$\mathbf{r} = \frac{Sxy}{\sqrt{Sxx} * Syy}$$

$$r = \frac{-7.749}{\sqrt{1730,90*35.312}}$$

r = - 0,991171 = - 99,1%

l) O coeficiente de determinação

$r^2 = -0,991171^2$
$r^2 = 0,982419 = 98,2$ %

m) O coeficiente de determinação ajustado

$$\text{‘}r^2 = 1 - \left(\frac{n-1}{n-2}\right) * (1 - r^2)$$

$$\text{‘}r^2 = 1 - \left(\frac{10-1}{10-2}\right) * (1 - 0,982419)$$

$\text{‘}r^2 = 0,9802 = 98,02\%$

Figura 38: Atividade 7 resolvida pelo aluno e enviada via email - parte 3

n) Os testes de hipóteses *t* e o intervalo de confiança de 95% para os parâmetros β_0 e β_1.

- Para β_0

$$\begin{cases} H_o = \beta_0 = 0 \\ H_1 = \beta_0 \neq 0 \end{cases}$$

$$T_o = \frac{\beta 0}{6\beta 0} = \frac{511{,}16}{126{,}83} = 4{,}0303$$

[β0 - (t*6β0) ; β0 + (t * 6β0)]
[511,16 – (2,3060 * 126,832 ; 511,16 + (2,3060 * 126,83)]
[218,69 ; 803,63 777,62]

- Para β_1

$$\begin{cases} H_o = \beta_1 = 0 \\ H_1 = \beta_1 \neq 0 \end{cases}$$

$$T_o = \frac{\beta 1}{6\beta 1} = \frac{-4{,}4769}{0{,}21168} = -\ 21{,}494$$

[β1 - (t*6β1) ; β1 + (t * 6β1)]
[-4,4769 – (2,3060 * 0,21168) ; -4,4769 + (2,3060 * 0,21168)]
[-4,965 ; -3,9888]

Figura 39: Atividade 7 resolvida pelo aluno e enviada via email - parte 4

2 ATIVIDADE 8 RESOLVIDA PELO ALUNO

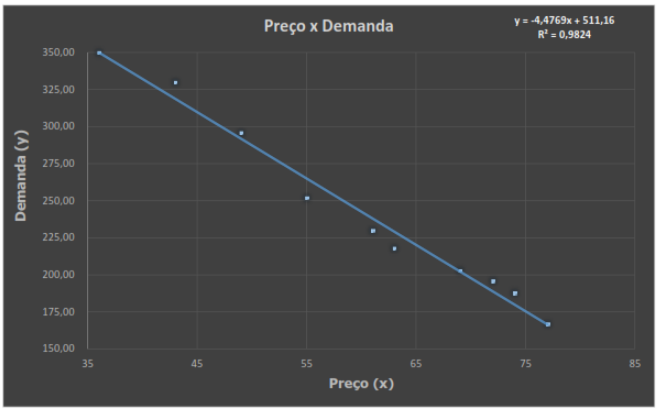
ITB – PROF. ANTÔNIO ARANTES FILHO			
Aluno: Jenilson Melo Nascimento		**Número:** 14	
3ª série:	**Curso:** Contabilidade	**Turma:** C	
Professor: Ricardo Will	**Disciplina:** Física	**Período:** Manhã	
	3 º TRIMESTRE	**Data:** 20 / set / 2016	**Nota**

Atividade 7 – Modelo de Regressão Linear Simples

02. Uma empresa resolveu estudar a variação da demanda (em unidades) de seu produto em função do preço de venda (em reais) praticado. Para isso, foram coletados os seguintes dados:

Tabela 16: Variação da demanda do produto em função do preço de venda

Preço (X)	Demanda (Y)
36	350,00
43	330,00
49	296,00
55	252,00
61	230,00
63	218,00
69	203,00
72	196,00
74	188,00
77	167,00



Determine:

a) Os estimadores Sxx, Sxy, Syy

	Preço (X)	Demanda (Y)	(Xi - X)²	(Yi - Y)²	(Xi - X)(Yi - Y)
	36	350,00	571,21	11.449,00	- 2.557,30
	43	330,00	285,61	7.569,00	- 1.470,30
	49	296,00	118,81	2.809,00	- 577,70
	55	252,00	24,01	81,00	- 44,10
	61	230,00	1,21	169,00	- 14,30
	63	218,00	9,61	625,00	- 77,50
	69	203,00	82,81	1.600,00	- 364,00
	72	196,00	146,41	2.209,00	- 568,70
	74	188,00	198,81	3.025,00	- 775,50
	77	167,00	292,41	5.776,00	- 1.299,60
Totais	599,00	2.430,00	1.730,90	35.312,00	- 7.749,00
Médias	59,90	243,00			

$S_{xx} = \sum(X_i - X)^2$
$S_{xx} = 1.730,90$

$S_{yy} = \sum(Y_i - Y)^2$
$S_{yy} = 35.312,00$

$S_{xy} = \sum(X_i - X)(Y_i - Y)$
$S_{xy} = -7.749,00$

b) Os parâmetros β_o e β_1.

$$\beta_1 = \frac{S_{xy}}{S_{xx}} = \frac{-7.749,00}{1.730,90} = -4,4769$$

$$\beta_0 = Y - \beta_1 X$$
$$\beta_0 = 243 - (-4,4769)(59,90)$$
$$\beta_0 = 511,16$$

Figura 40: Atividade 8 resolvida pelo aluno e enviada via email - parte 1

c) A equação de regressão linear simples

$Y = -4{,}4769x + 511{,}16$

d) A soma dos quadrados dos resíduos (SQR)

$\mathbf{SQR = Syy - \beta_1 * Sxy}$

$SQR = 35.312 - (-4{,}4769 * -7.749)$

$SQR = 620{,}50$

e) O quadrado médio dos resíduos (QMR)

$\mathbf{QMR} = 6^2$

$\mathbf{QMR} = \frac{SQR}{n-2}$

$QMR = \frac{620{,}50}{8}$

$QMR = 77{,}56$

f) O desvio padrão dos resíduos

$\sigma = \sqrt{6^2}$

$\sigma = \sqrt{77{,}56}$

$\sigma = 8{,}8068$

g) A variância de β_0 e β_1

$\mathbf{var\ (\beta_0)} = \sqrt{\frac{6^2}{\beta 0}}$

$\mathbf{var\ (\beta_0) = \sigma^2 * \frac{1}{n} + \frac{X^2}{Sxx}}$

$var\ (\beta_0) = 77{,}56 * (\frac{1}{10} + \frac{(599)^2}{1730{,}90})$

$var\ (\beta_0) = 16.085{,}29$

$\mathbf{var\ (\beta_1)} = \sqrt{\frac{6^2}{\beta 1}}$

$\mathbf{var\ (\beta_1)} = \frac{6^2}{Sxx}$

$var\ (\beta_1) = \frac{77{,}56}{1730{,}90}$

$var\ (\beta_1) = 0{,}04481$

h) O desvio padrão de β_0 e β_1

$\mathbf{\sigma_{\beta 0}} = \sqrt{\frac{6^2}{\beta 0}}$

$\sigma_{\beta 0} = \sqrt{16.085{,}29}$

$\sigma_{\beta 0} = 126{,}83$

$\mathbf{\sigma_{\beta 1}} = \sqrt{\frac{6^2}{\beta 1}}$

$\sigma_{\beta 1} = \sqrt{0{,}04481}$

$\sigma_{\beta 1} = 0{,}21168$

Figura 41: Atividade 8 resolvida pelo aluno e enviada via email - parte 2

i) A análise da variância (ANOVA) completando a tabela

ANOVA

	gl	*SQ*	*MQ*	*F*	*F de significação*
Regressão	1	34691,20169	34691,20169	447,052783	2,63056E-08
Resíduo	8	620,798313	77,59978913		
Total	9	35312			

SQT = Syy
SQT = 35.312

SQE = QME
SQE = SQT – SQR
SQE = 35.312 – 620,50
SQE = 34.691,50

j) O teste F para um nível de confiança de 95%

$$F_0 = \frac{QME}{QMR} \qquad F_0 = \frac{34.691,5}{77,56}$$

$$F_0 = 447,29$$

k) O coeficiente de correlação Pearson.

$$\mathbf{r = \frac{Sxy}{\sqrt{Sxx * Syy}}}$$

$$r = \frac{-7.749}{\sqrt{1730,90 * 35.312}}$$

r = - 0,991171 = - 99,1%

l) O coeficiente de determinação

$r^2 = -0,991171^2$
$r^2 = 0,982419 = 98,2\ \%$

m) O coeficiente de determinação ajustado

$$\mathbf{‘r^2 = 1 - (\frac{n-1}{n-2}) * (1 - r^2)}$$

$$‘r^2 = 1 - (\frac{10-1}{10-2}) * (1 - 0,982419)$$

$‘r^2 = 0,9802 = 98,02\%$

Figura 42: Atividade 8 resolvida pelo aluno e enviada via email - parte 3

n) Os testes de hipóteses *t* e o intervalo de confiança de 95% para os parâmetros β_0 e β_1.

- Para β_0

$$\begin{cases} H_o = \beta_0 = 0 \\ H_1 = \beta_0 \neq 0 \end{cases}$$

$$T_o = \frac{\beta 0}{6\beta 0} = \frac{511{,}16}{126{,}83} = 4{,}0303$$

$$[\beta 0 - (t * 6\beta 0) \quad ; \quad \beta 0 + (t * 6\beta 0)]$$
$$[511{,}16 - (\; 2{,}3060 * 126{,}832 \quad ; \quad 511{,}16 + (2{,}3060 * 126{,}83)]$$
$$[218{,}69 \quad ; \; 803{,}63 \quad 777{,}62]$$

- Para β_1

$$\begin{cases} H_o = \beta_1 = 0 \\ H_1 = \beta_1 \neq 0 \end{cases}$$

$$T_o = \frac{\beta 1}{6\beta 1} = \frac{-4{,}4769}{0{,}21168} = -\,21{,}494$$

$$[\beta 1 - (t * 6\beta 1) \quad ; \quad \beta 1 + (t * 6\beta 1)]$$
$$[-4{,}4769 - (2{,}3060 * 0{,}21168) \quad ; \quad -4{,}4769 + (2{,}3060 * 0{,}21168)]$$
$$[-4{,}965 \quad ; \; -3{,}9888]$$

Figura 43: Atividade 8 resolvida pelo aluno e enviada via email - parte 4

BIBLIOGRAFIA

[1] ALLARD, R. J. **Introdução á Econometria**. Rio de Janeiro: Ed. 1980.

[2] BARBANCHO, Alfonso G. Alfonso G. **Complementos de econometría**. 1977.

[3] BULMER, Michael. **Francis Galton: pioneer of heredity and biometry**. London: Johns Hopkins University Press, 2003.

[4] BARBETTA, P.A.: **Estatística aplicada às Ciências Sociais**. Florianópolis. Editora da UFSC, 2010, Tabela 5.

[5] CHRIST, Carl F. **Econometric models and methods**. 1966.

[6] **Economic Report of the President**. feb-2015, Table B-2, p. 386.

[7] **Economic Report of the President**. feb-2016, Table B-2, p. 406.

[8] Bryant, Bradley J. **Como calcular PMC e PMP**. eHOW Brasil, http://www.ehow.com.br/calcular-pmc-pmp-como_75149/.

[9] Mcbride, Carter. **Como calcular a propensão marginal a poupar**. eHOW Brasil, http://www.ehow.com.br/calcular-propensao-marginal-poupar-como_293124/.

[10] FALCO, Javert Guimarães. **Estatística aplicada**, EdUFMT; Curitiba: UFPR, 2008

[11] FIGUEIREDO FILHO, Dalson Brito; SILVA JUNIOR, José Alexandre. **Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r)**. Revista Política Hoje, v. 18, n. 1, 2010.

[12] GALTON, F. **Hereditary Talent and Character**. In: MacMillan's Magazine, London, vol. XII, p. 157-166 / p. 318-327, junho-agosto de 1865.

[13] GRINGS, José Fernando. **Estimação por Intervalo de Confiança da Média aula 16**. https://www.youtube.com/watch?v=4t2OsjQwlcI

[14] GUJARATI, Damodar N.; PORTER, Dawn C. **Econometria Básica**-5. AMGH Editora, 2011.

[15] HILL, R. Carter; GRIFFITHS, William E.; JUDGE, George G. **Econometria**. Saraiva, 2003.

[16] HOFFMANN, Rodolfo et al. **Análise de regressão: uma introdução à econometria**. 2015.

[17] **Introdução à Economia**, UnB. https://www.introducaoaeconomia.files.wordpress.com/2015/05/gabarito-lista-5b1.pdf

[18] KHAN Academy, **Teste de hipótese com amostra pequena**, https://www.youtube.com/watch?v=TJbnkmiZiRU&feature=youtu.be

[19] KEYNES, John Maynard. **General theory of employment, interest and money**. Atlantic Publishers e Dist, 2006.

[20] INVESTOPEDIA, **Ceteris paribus**. http://www.investopedia.com/terms/c/ceteris-paribus.asp

[21] INVESTOPEDIA, **Regressão**. http://www.investopedia.com/video/play/regression/

[22] LANGE, Oskar. **Introdução à econometria**. Fundo de Cultura, 1967.

[23] MATOS, Orlando Carneiro. **Econometria básica: teoria e aplicações**. Atlas, 1995.

[24] MILONE, Giuseppe; ANGELINI, Flávio. **Estatística aplicada: números-índices, regressão e correlação, séries temporais**. São Paulo: Atlas, 1995.

[25] MORETTIN, Pedro Alberto. **Estatística Básica - Probabilidade**. São Paulo: Makros Books, 1999.

[26] MOTTA, Regis Da Rocha et al. **Engenharia econômica e finanças**. Elsevier Brasil, 2009.

[27] PSICOSABER. **Francis Galton**. https://www.psicosaber.wordpress.com/2009/05/18/francis-galton/

[28] Portal Action. **Estimação dos parâmetros do modelo**, http://www.portalaction.com.br/analise-de-regressao/12-estimacao-dos-parametros-do-modelo

[29] Portal Action. **Análise de variância**, http://www.portalaction.com.br/analise-de-regressao/15-analise-de-variancia

[30] **Regressão Linear Simples**, UNICAMP. http://www.ime.unicamp.br/ hlachos/regresslide.pdf

[31] Silva, Carla da, **Teste de hipótese da média para sigma conhecido**. https://www.youtube.com/watch?v=fQO1lje8kKY

[32] Suporte ao Minitab. **O que é ANOVA**, http://support.minitab.com/pt-br/minitab/17/topic-library/modeling-statistics/anova/basics/what-is-anova/

[33] SURREY, Michael James Charles. **An introduction to econometrics**. Clarendon Press, 1979.

[34] TINBERGEN, Jan. Annual survey: **Suggestions on quantitative business cycle theory**. Econometrica, Journal of the Econometric Society, p. 241-308, 1935.

[35] TINTNER, Gerhard. **Elementos de econometria**. Pioneira, 1965.

[36] VARTANIAN, P. R.; CIA, J. C.; MENDES-DA-SILVA, W. **Econometria: análise de dados com regressão linear**. Säo Paulo: Saint Paul, 2013.

[37] FERNANDES, T. **Keynes não defendia estado forte, destaca pesquisador**, http://www.http://veja.abril.com.br/noticia/economia/keynes-nao-defendia-estado-forte-destaca-pesquisador.

[38] WOOLDRIDGE, Jeffrey M. **Introdução à Econometria: Uma Abordagem Moderna**. Cengage Learning, 2015.