

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

ESTIMAÇÃO DE INDICADORES DE PROJETO PARA  
ALVENARIA ESTRUTURAL

Ricardo Blanco

Trabalho de Conclusão de Curso

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

ESTIMAÇÃO DE INDICADORES DE PROJETO PARA  
ALVENARIA ESTRUTURAL

Ricardo Blanco

Orientador: Profa. Dra. Teresa Cristina Martins Dias

Co-Orientador: Prof. Dr. Guilherme Aris Parsekian

Este exemplar trata da  
redação do Trabalho de  
Conclusão de Curso a ser  
devidamente corrigido e  
apresentado por Ricardo  
Blanco e a ser avaliada pela  
comissão julgadora

São Carlos, 19 de junho de 2009

Trabalho de Conclusão de Curso

## **RESUMO**

Este trabalho trata da construção de modelos que visam estimar diversos indicadores de projetos, como taxa de aço em alvenaria por área do pavimento tipo e de identificar quais variáveis, como índice de compactação ou presença de pilotis, influenciam estes indicadores, como por exemplo, a taxa de aço em alvenaria por área do pavimento tipo. Para o estudo da estimação de tais ajustes pode-se utilizar técnicas de análise de regressão e análise multivariada.

Há ainda a ressaltar que um estudo sobre indicadores de projetos em alvenaria estrutural está sendo desenvolvido pelo aluno Guilherme Martins Lopes, orientado pelo professor Doutor Guilherme Aris Parsekian (departamento de Engenharia Civil – DECiv/UFSCar).

**Palavras chave:** *Análise de Regressão, Análise Multivariada, Alvenaria Estrutural*

# SUMÁRIO

<b>CAPÍTULO 1.....</b>	<b>1</b>
<b>1.1. JUSTIFICATIVA.....</b>	<b>2</b>
<b>1.2. OBJETIVO .....</b>	<b>2</b>
<b>1.3. DESCRIÇÃO DOS CAPÍTULOS .....</b>	<b>3</b>
<b>CAPÍTULO 2.....</b>	<b>5</b>
<b>2.1. INDICADORES UTILIZADOS .....</b>	<b>5</b>
<b>2.2. VARIÁVEIS QUALITATIVAS UTILIZADAS .....</b>	<b>6</b>
<b>2.3. DESCRIÇÃO DOS DADOS UTILIZADOS .....</b>	<b>7</b>
<b>CAPÍTULO 3.....</b>	<b>9</b>
<b>3.1. REGRESSÃO SIMPLES .....</b>	<b>9</b>
<b>3.2.1. MÉTODO DOS MÍNIMOS QUADRADOS .....</b>	<b>11</b>
<b>3.3.1. MODELO DE REGRESSÃO MÚLTIPLA.....</b>	<b>13</b>
<b>3.3.2. ESTIMAÇÃO DOS COEFICIENTES DE REGRESSÃO.....</b>	<b>14</b>
<b>3.3.3. CÁLCULO DOS RESÍDUOS .....</b>	<b>15</b>
<b>3.3.4. TESTE F.....</b>	<b>17</b>
<b>3.3.5. COEFICIENTE DE DETERMINAÇÃO DO MODELO.....</b>	<b>18</b>
<b>3.3.6. COEFICIENTE DE VARIAÇÃO DO MODELO .....</b>	<b>19</b>
<b>3.4. ANÁLISE DE CLUSTER OU ANÁLISE DE AGRUPAMENTOS.....</b>	<b>19</b>
<b>3.5. CONSTRUÇÃO DOS GRUPOS.....</b>	<b>20</b>
<b>3.6. MÉTODOS HIERÁRQUICOS .....</b>	<b>20</b>
<b>3.7. VERIFICAÇÃO DA HETEROGENEIDADE ENTRE GRUPOS.....</b>	<b>21</b>
<b>3.8. IDENTIFICANDO O NÚMERO DE GRUPOS .....</b>	<b>21</b>
<b>CAPÍTULO 4.....</b>	<b>23</b>
<b>4.1. CONSIDERAÇÕES FINAIS .....</b>	<b>23</b>

# Capítulo 1

## Considerações iniciais

A estatística, que durante muito tempo foi uma ciência praticamente desconhecida para grande parte das pessoas e que para muitos daqueles que à “conheciam” era apenas utilizada para cálculos de censo e afins, tem sido amplamente reconhecida por diversos pesquisadores e suas técnicas aplicadas em várias áreas do conhecimento. Assim, devido ao grande número de áreas que a ciência estatística está relacionada, é essencial que tais profissionais tenham alta capacidade de comunicação, dado que quase sempre terá que atuar com pessoas de outros campos.

Como exemplo, este trabalho será aplicado na área da construção civil, em que técnicas estatísticas podem ser utilizadas para a melhoria e otimização de seus procedimentos, tais como a estimação de quantidade de concreto, aço e outras variáveis necessárias em uma construção.

Há ainda a destacar, que a construção civil tem grande importância econômica e social, ajudando na formação do PIB e geração de empregos. Um exemplo de tal importância pode ser observado com o início da atual crise econômica mundial, em que várias das demissões que inicialmente ocorreram nos EUA foram no setor da construção civil.

O setor da construção civil pode ser dividido em três subsetores:

- Construção pesada: relacionada a obras de infra-estrutura (vias, saneamento, hidrelétrica, etc);

- Construção industrial: relacionada à montagem de indústrias, telecomunicações, geração e transmissão de energia, etc;
  - Construção de edificações: relacionada ao planejamento, construção e reforma de edifícios (residenciais, comerciais, institucionais, etc)
- (FUNDAÇÃO JOÃO PINHEIRO, 1984([1]); FARAH, 1992([2]))

## **1.1. Justificativa**

A construção civil é uma ciência que geralmente trabalha com grandes quantidades de material (como aço, concreto, argamassa, etc.) e está presente em todos os países. Assim, a estimação de indicadores que possam ajudar a otimizar o gasto com material e tempo em uma obra é de grande importância. Há ainda a destacar que com a disponibilização de tais indicadores o engenheiro pode analisar mais precisamente seu projeto, identificando possíveis erros e/ou melhorando. Quanto à minimização de custos o impacto se dá na geração de novos empregos, sendo que construções mais baratas (contemplando os padrões mínimos de segurança requeridos) implicam em maior facilidade de compra e conseqüentemente maior demanda pelos produtos, gerando mais empregos e poder de compra para determinados setores da população.

## **1.2. Objetivo**

Este trabalho tem como objetivo principal a estimação de indicadores utilizados na construção civil, utilizando modelos de regressão. Pretende se construir modelos específicos para determinados grupos (baseados nas tipologias arquitetônicas, ou seja, número de andares, forma arquitetônica etc) que poderão ser criados com o uso de

técnicas de análise multivariada. Há ainda a destacar que tais decisões devem ser amparadas pelas informações, indiscutivelmente essenciais, dos pesquisadores da área.

### **1.3. Descrição dos Capítulos**

Este trabalho se divide em 4 capítulos seguido de uma bibliografia. Os capítulos em questão são:

Capítulo 1: Apresentação de uma introdução ao problema a ser estudado assim como a justificativa da necessidade do trabalho e objetivo do mesmo;

Capítulo 2: Apresentação das variáveis quantitativas a serem utilizadas, e descrição dos dados;

Capítulo 3: Metodologia que se pretende utilizar para a resolução do problema proposto.

Capítulo 4: Propostas futuras



## Capítulo 2

### O projeto em alvenaria estrutural e algumas especificações

Os indicadores usados neste trabalho foram obtidos de vários projetos de diversas tipologias arquitetônicas, que englobam aspectos como altura do edifício e arquitetura do mesmo, sendo que cada tipologia apresenta consumo de materiais diferenciados. Tais indicadores, podem ser utilizados para identificar aspectos no projeto a serem otimizados visando a diminuição de custo, sem o comprometimento da obra.

Projetos em alvenaria tratam da construção de estruturas e paredes utilizando unidades conectadas entre si por meio de argamassa. O termo alvenaria estrutural é utilizado quando se trata da estrutura de uma edificação.

#### **2.1. Indicadores utilizados**

Os indicadores que são descritos a seguir tratam de covariáveis que inicialmente estão presentes nos ajustes de todos os modelos, somente sendo retiradas quando não existir significância, valendo destacar que os únicos ajustes que não contêm todas as variáveis descritas abaixo são os sobrados, visto que não possuem Índice de Circulação.

As covariáveis quantitativas a serem utilizadas nos ajustes são:

- Índice de Circulação: referente à área de circulação de uma edificação, ou seja, a área que é destinada à circulação das pessoas, onde do ponto de vista financeiro é bastante cara, visto que não gera lucro algum;
- Índice de Compacidade: indicação do quanto a construção se distancia da estrutura mais econômica em relação ao perímetro das paredes externas;
- Densidade das Paredes: área ocupada pelas paredes com relação à área total da construção.
- Relação da área de alvenaria estrutural por alvenaria total: Razão entre a área de paredes estruturais por área total.

## 2.2. Variáveis qualitativas utilizadas

Para a estimação dos indicadores, além das variáveis quantitativas descritas na seção 2.1, são utilizadas variáveis que entram nos modelos como variáveis *dummies*.

As variáveis utilizadas são:

- Presença de Cinta meia altura-paredes externas
- Presença de Cinta meia altura paredes externas e internas
- Presença de Grautes verticais encontros de parede
- Presença de laje maciça
- Presença de laje pré moldada com vigotas
- Presença de amarradura solta
- Presença de amarradura em tela
- Presença de Fundação Estaca
- Presença de Fundação Tubulão

Para maiores detalhes sobre as variáveis citadas ver Freire, Ailton Soares, 2008 ([3]).

## 2.3. Descrição dos dados utilizados

O banco de dados utilizado no desenvolvimento deste trabalho, é formado de 107 observações e 55 variáveis, das quais 29 são definidas como variáveis resposta e 13 como covariáveis. Há ainda a destacar que das covariáveis analisadas apenas **índice de compacidade, densidade das paredes, índice de circulação e relação da área de alvenaria de estrutural por alvenaria total** são quantitativas e utilizadas no ajuste de todos os modelos, sendo as outras qualitativas e somente presentes em alguns casos específicos.

Tabela 2.1: Estatísticas descritivas das variáveis quantitativas.

Variável	Média	Desvio Padrão	Mínimo	Máximo	Coefficiente de Variação
<b>Índice de Compacidade</b>	65,4880	10,6138	28,0100	89,0700	16,2073
<b>Densidade das Paredes</b>	9,7594	1,1583	5,5400	12,1700	11,8681
<b>Índice de Circulação</b>	9,9385	5,3763	0,0000	29,6400	54,0953
<b>Relação da área de alvenaria estrutural por alvenaria total</b>	93,2234	5,3823	70,6504	98,8946	5,7735

Os conjuntos de ajustes mostrados na Tabela 2.2 são formados por vários modelos, que diferem pelas particularidades, como por exemplo a presença de pilotis. Esta tabela foi construída a partir da tabela original(Tabela 10, página 34) de Martins[4].

Tabela 2.2: Conjuntos de análises realizadas.

<b>Conjunto de Análises</b>	<b>O que se espera medir</b>
<b>Conjunto 1</b>	Influência geral dos parâmetros de arquitetura na alvenaria estrutural
<b>Conjunto 2</b>	Influência de cintas no consumo de graute e armadura em alvenaria
<b>Conjunto 3</b>	Influência do tipo de laje e armadura na espessura da laje e consumo de aço
<b>Conjunto 4</b>	Influência do tipo de fundação no número de elementos de fundação
<b>Conjunto 5</b>	Influência geral dos parâmetros de arquitetura nos consumos do pilotis

## Capítulo 3

### Métodos utilizados

O trabalho em questão utiliza as metodologias de análise de regressão e análise de *cluster*, que são descritas neste capítulo

#### 3.1. Regressão Simples

Análise de regressão é uma metodologia estatística que estuda a relação entre duas ou mais variáveis. O estudo da relação entre duas variáveis é denominado de regressão simples e o estudo da relação entre mais de duas é chamado de regressão múltipla.

Em regressão simples, tem-se uma variável resposta, ou seja, uma variável dependente (Y) e uma covariável (X), independente de Y.

Uma das maneiras mais fáceis de identificar a existência de uma relação linear passível do ajuste de um modelo de regressão simples é a construção de um diagrama de pontos sendo possível observar a existência de correlação entre as variáveis estudadas. Como exemplo, o Gráfico 2.1 mostra a situação hipotética em que a taxa de aço em alvenaria por área do pavimento tipo tem seu número aumentado quanto maior a densidade das paredes. Este gráfico foi construído no software Minitab ([10]). Assim temos:

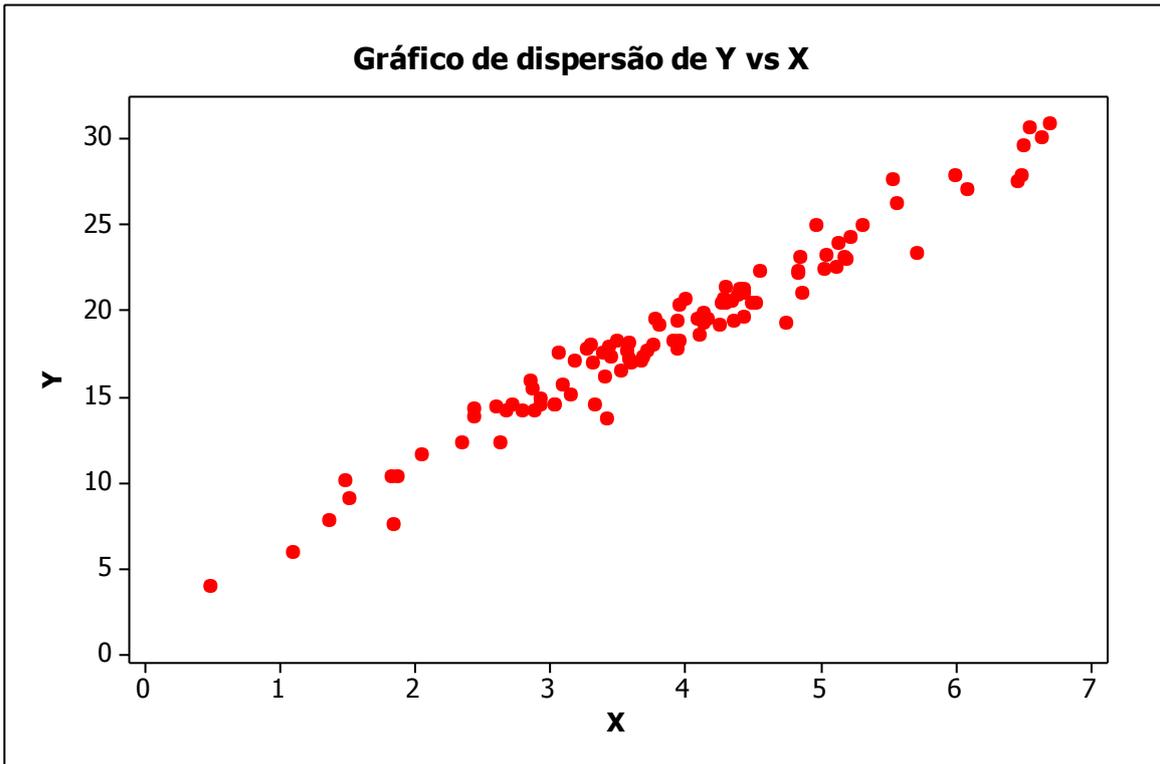


Gráfico 3.1: Exemplo de relação linear entre X e Y.

Observando o gráfico de dispersão (Gráfico 2.1), percebe-se claramente a relação linear entre a variável Y (taxa de aço em alvenaria por área do pavimento tipo) e X (Densidade das paredes).

A partir dos dados observados é possível ajustar uma reta que passe pelos pontos de forma que os erros sejam minimizados. Assim, a estimação dos parâmetros da reta ajustada é dada por

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon, \quad (3.1)$$

sendo que para  $i=1, \dots, n$

$Y_i$  = Valor da resposta para a  $i^{\text{th}}$  observação

$\beta_0$  = Coeficiente linear da reta

$\beta_1$  = Coeficiente angular da reta

$X_i$  =  $i^{\text{th}}$  valor observado para a covariável X

$\varepsilon_i$  =  $i^{\text{th}}$  valor observado de um vetor com média 0 e variância  $\sigma^2$

Por fim, quando é realizado o ajuste de um modelo de regressão para estimar a taxa de aço em alvenaria por área do pavimento tipo com a densidade das paredes como covariável, são obtidos os resultados:

$$\text{Coeficiente Linear}(\beta_0) = 2,8888$$

$$\text{Coeficiente Angular}(\beta_1) = 4,02743$$

Como interpretação da reta ajustada,  $\hat{Y}_i = 2.8888 + 4.02743X_i$ , cada aumento de uma unidade em X, representa adição de 4,02743 unidades em Y. No Gráfico 2.2, é mostrado o ajuste da reta.

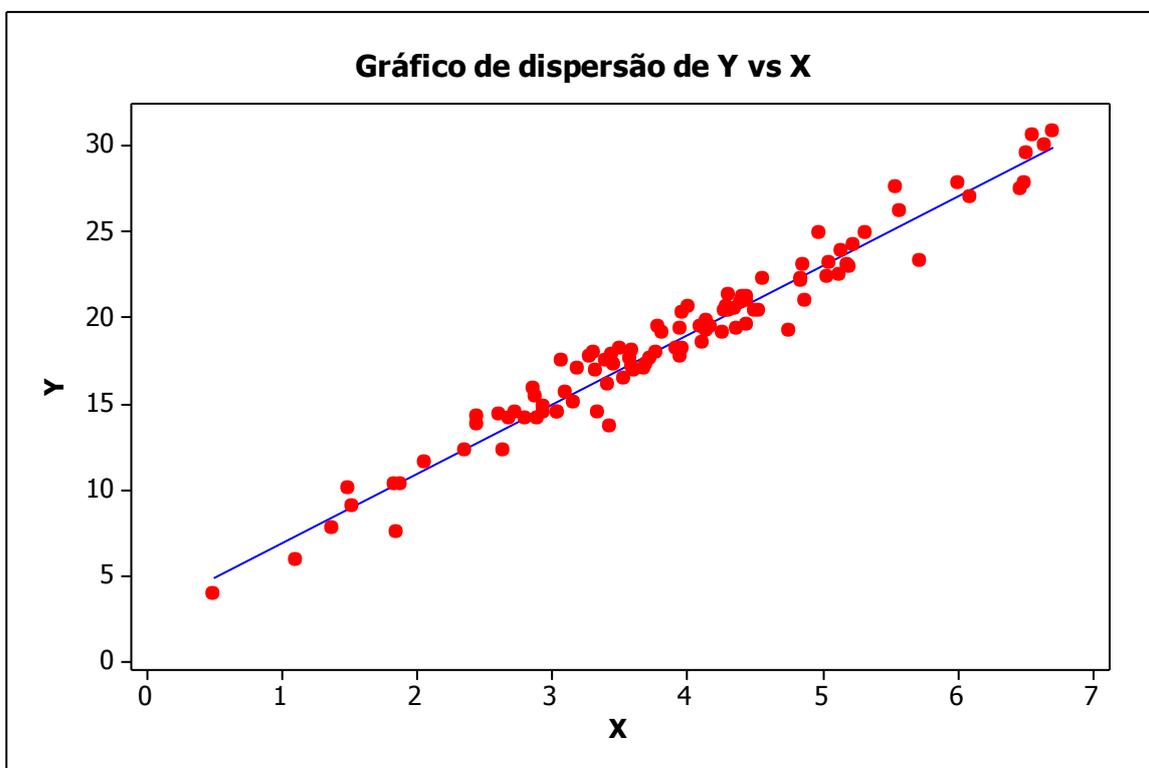


Gráfico 3.2: Reta de regressão ajustada para Y como função de X

### 3.2.1. Método dos Mínimos Quadrados

Para encontrar os estimadores dos parâmetros do modelo de regressão simples, foi utilizado o método dos mínimos quadrados, que se baseia na tentativa de minimizar

o quadrado da diferença existente entre o valor observado e o estimado, ou seja, minimizar o quadrado dos erros, isto é,

$$Q = \sum_{i=1}^n (Y_i - \beta_0 + \beta_1 X_i)^2 \quad (3.2)$$

As estimativas são encontradas a partir da resolução do sistema de equações, mostrado nas equações 2.3.

$$\begin{aligned} \frac{\partial Q}{\partial \beta_0} &= -2 \sum (Y_i - \beta_0 - \beta_1 X_i) = 0 \\ \frac{\partial Q}{\partial \beta_1} &= -2 \sum X_i (Y_i - \beta_0 - \beta_1 X_i) = 0 \end{aligned} \quad (3.3)$$

A resolução fornece os estimadores dados em 2.4.

$$\begin{aligned} \beta_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \beta_1 &= \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned} \quad (3.4)$$

Para maior detalhes ver, NETER, John [5].

### 3.3. Regressão Múltipla

A regressão múltipla é uma extensão dos modelos de regressão linear simples.

Suponha que se deseja observar a quantidade de aço que será utilizada em um prédio, e para isso foi observado a altura e área da construção. Assim, no modelo de regressão múltipla, a variável resposta (variável dependente) é a quantidade de aço a ser utilizada e as variáveis preditoras (variáveis independentes) são a área e a altura.

Há ainda a destacar que para a construção de um modelo com as seguintes características, deve-se ter em mãos uma amostra aleatória de observações sobre prédios

que possuam as variáveis: quantidade de aço, área e altura. A importância do uso desta metodologia no problema a ser trabalhado, é a possibilidade de realizar estimações confiáveis de indicadores utilizados na construção civil, tais como quantidade de aço e argamassa a serem utilizadas.

Nos casos em que se trabalha com modelos de regressão múltipla, um procedimento bastante utilizado para a escolha das covariáveis que possuam relação significativa com a variável resposta, é a observação do coeficiente de correlação, em que valores baixos indicam a falta de informação obtida naquela variável com relação à resposta a ser estimada, análise do VIF (*Variance Inflation Factor*), que visa selecionar variáveis com ausência de multicolinearidade, ou seja, as variáveis de entrada possuem relações no mínimo aproximadamente lineares entre si, o que implica em variáveis diferentes estarem explicando praticamente as mesmas coisas, e por fim, a utilização de algoritmos de seleção de variáveis, como o *Stepwise*, *Backward* e *Forward*, (NETER, John [5]).

### 3.3.1. Modelo de Regressão Múltipla

Considere a existência de  $p-1$  variáveis preditoras. Sendo assim, temos que o modelo de regressão geral é dado por:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i(p-1)} + \varepsilon_i \quad (3.5)$$

Assumindo que  $\varepsilon_i$  possui média igual a zero, temos:

$$E(Y)_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i(p-1)} \quad (3.6)$$

sendo:

$\beta_i =$  vetor de parâmetros para  $i = 0, \dots, p-1$

$X_{1i}, \dots, X_{i,(p-1)} =$  Co variáveis

$\varepsilon_i =$  erros independentes com  $N \sim (0, \sigma^2 I), i = 1, \dots, n$

$p =$  número de covariáveis

$n =$  número de observações

De forma matricial, tem-se que:

$$Y_i = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_{n-1} \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdot & \cdot & \cdot & X_{1(p-1)} \\ 1 & X_{21} & X_{22} & \cdot & \cdot & \cdot & X_{2(p-1)} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & X_{(n-1)1} & X_{n-12} & \cdot & \cdot & \cdot & X_{(n-1)(p-1)} \\ 1 & X_{n1} & X_{n2} & \cdot & \cdot & \cdot & X_{n(p-1)} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_{p-2} \\ \beta_{p-1} \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_{n-1} \\ \varepsilon_n \end{bmatrix} \quad (3.7)$$

Por fim, o modelo de regressão linear pode ser escrito como:

$$Y = X\beta + \varepsilon \quad (3.8)$$

### 3.3.2. Estimação dos Coeficientes de Regressão

Considerando que interesse reside na estimação dos coeficientes de regressão, e aplicando a esperança na expressão (3.8) tem-se que:

$$E(Y) = E(X)\beta + E(\varepsilon) \quad (3.9)$$

Assumindo que os erros seguem distribuição Normal( $0, \sigma^2$ ), obtém-se:

$$Y = X\beta \quad (3.10)$$

Aplicando algumas operações algébricas em (3.10), o vetor estimado de parâmetros  $\beta$  é dado por:

$$\begin{aligned}
X'Y &= X'X\beta \\
(X'X)^{-1}X'Y &= (X'X)^{-1}X'X\beta \\
(X'X)^{-1}X'Y &= I\beta \\
\boxed{\hat{\beta} &= (X'X)^{-1}X'Y}
\end{aligned}
\tag{3.11}$$

E portanto, os valores estimados para Y são dados por:

$$\hat{Y} = (X(X'X)^{-1}X'Y), \tag{3.12}$$

que é equivalente a  $\hat{Y} = HY$  quando e  $X(X'X)^{-1}X'$  é substituído por **H**.

### 3.3.3. Cálculo dos Resíduos

Sejam o vetor de valores ajustados  $\hat{Y}$  dado em (3.12). O vetor de resíduos é dado por  $e_i = Y_i - \hat{Y}_i$ , podendo ser escrito na forma  $e_i = (I - H)Y$ . Vale ressaltar aqui a suposição de que os mesmos seguem distribuição Normal com vetor de médias igual a zero e matriz de variância-covariância igual a  $\sigma^2(I - H)$ , sendo estimada por  $MSE(I - H)$ .

Um fator importante relacionado aos resíduos é a possibilidade de se verificar a existência de pontos atípicos ou algum tipo de tendência nos dados, como por exemplo, a necessidade da adição de um termo quadrático no modelo. Destaca-se ainda que é extremamente importante que os resíduos finais, obtidos após o ajuste de cada modelo, devem ter distribuição normal com média zero e variância constante, já que todos os cálculos tomam por base tais suposições. Caso a normalidade e homocedasticidade não sejam aceitas pode-se utilizar transformações, como as do tipo Box-Cox ou o uso de modelos heterocedásticos (Cribari Neto,[1]).

Retornando ao exemplo descrito na seção 3.1, a análise de resíduos para o conjunto em questão é mostrada nos Gráficos 3.3, sendo que os resultados logo abaixo servem como modelo de um bom ajuste.

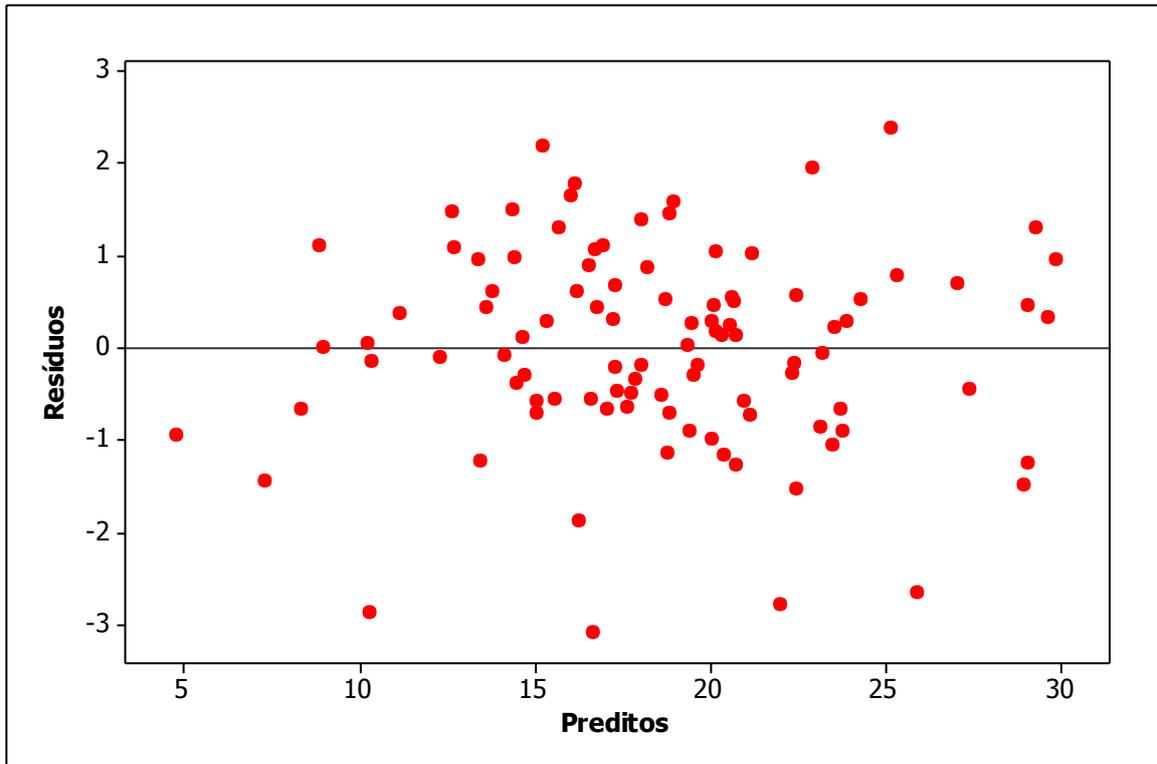


Gráfico 3.3: Resíduos x Preditos

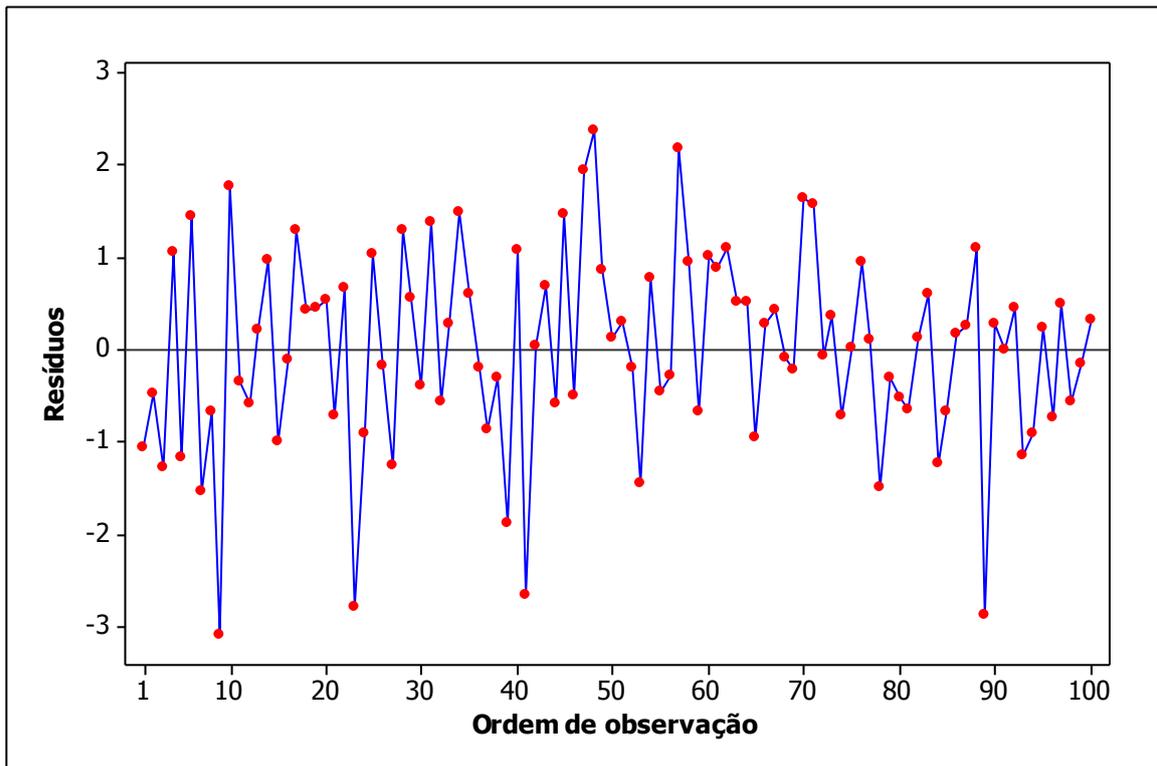


Figura 3.4: Resíduos x Ordem

Tabela 3.1: Teste de Normalidade.

<b>Teste de Normalidade</b>		
<b>Teste</b>	<b>Estatística</b>	<b>P - Valor</b>
<b>Cramer - von Mises</b>	0.034	> 0.25
<b>Anderson - Darling</b>	0.356	> 0.25

Assim, observando os resultados acima, percebemos homcedasticidade nos resíduos, ausência de tendência e normalidade dos mesmos.

### 3.3.4. Teste F

O teste F é utilizado para verificar a existência de relação entre a variável resposta e as covariáveis. As hipóteses definidas para este teste implicam em:

$H_0$ : nenhuma variável é significativa;

$H_1$ : pelo menos uma variável é significativa.

ou seja,

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$$

$$H_1: \beta_i \neq \beta_j, \text{ para ao menos um } i \neq j$$

sendo que  $F = \frac{MSR}{MSE} \sim F_{(1-\alpha; p-1, n-p)}$ ,  $MSR = \frac{Y' \left[ H - \left( \frac{1}{n} \right) J \right] Y}{p-1}$  e  $MSE = \frac{Y' [I - H] Y}{n-p}$

Se  $H_0$  é aceito, conclui-se que nenhuma variável utilizada no ajuste do modelo é significativa para explicar a resposta, caso contrário pelo menos uma das variáveis é, portanto, o modelo é significativo.

### 3.3.5. Coeficiente de Determinação do Modelo

O coeficiente de determinação do modelo, conhecido como  $R^2$ , é dado por:

$R^2 = \frac{SSR}{SST} * 100$ , em que SSR é a soma de quadrados da Regressão e SST a soma de quadrados total, sendo que  $0 \leq R^2 \leq 1$ ,

onde,

$$SSR = b' X' Y - \left( \frac{1}{n} \right) Y' J Y \quad (3.13)$$

$$SST = Y' Y - \left( \frac{1}{n} \right) Y' J Y, \quad (3.14)$$

sendo J igual à uma matriz quadrada de ordem n de 1's.

Este coeficiente, muitas vezes é erroneamente utilizado para verificar a adequabilidade do modelo ajustado, como mostrou Anscombe([2]). Portanto, tal estatística deve ser olhada cuidadosamente para que conclusões errôneas não sejam

tomadas. Vale destacar que neste trabalho, o coeficiente de determinação do modelo não é o único aspecto observado na análise de adequabilidade do modelo.

### 3.3.6. Coeficiente de Variação do Modelo

O coeficiente de variação trata de uma medida de dispersão que expressa a variabilidade de uma dada variável retirando a influência da ordem de grandeza da variável. Note que quanto menor o coeficiente mais homogêneo é o conjunto de dados. O coeficiente de variação é definido por,

$$CV = \frac{s}{\bar{x}} \quad (3.15)$$

ou ainda expresso na forma de porcentagem:

$$CV = \frac{s}{\bar{x}} 100\% \quad (3.16)$$

Destaca-se que a expressão (3.16) é bastante utilizada, visto que caracteriza a variabilidade dos dados em termos relativos à média. Outra característica importante do coeficiente de variação é que ele é adimensional e por isso pode ser utilizado para comparar dispersões de variáveis em unidades variadas.

## 3.4. Análise de Cluster ou Análise de Agrupamentos

A análise de *cluster* é um método estatístico que busca separar um determinado conjunto de observações em grupos homogêneos, de tal forma que a variabilidade dentro dos grupos seja menor do que entre os grupos, (Mardia ([8])). A importância do uso desta metodologia no problema em questão concentra-se no fato de que o uso de muitas tipologias implicaria em poucas observações para o ajuste de modelos de regressão, o que não seria tão interessante quanto um ajuste levando em consideração

várias observações com certo grau de homogeneidade. Assim, trabalhando com este método, pode-se buscar a redução do número de tipologias e maximizar a qualidade do ajuste de modelos de regressão múltipla.

Basicamente, a análise de *cluster* é composta por técnicas de agrupamento que buscam uma estrutura natural dos dados. Essa técnica não requer nenhuma suposição acerca do número de grupos ou suas estruturas, podendo tais grupos ser constituídos por observações individuais, multivariadas ou agrupamentos multivariados de variáveis que são analisadas, segundo as similaridades e dissimilaridades existentes entre eles.

### 3.5. Construção dos Grupos

Os grupos a serem construídos baseiam-se nas diferenças ou semelhanças entre os indivíduos em estudo. Assim, para avaliar tais similaridades é necessário a utilização de medidas de similaridades. Dentre as existentes, a utilizada neste trabalho é a Distância Euclidiana, definida a seguir e mostrada na equação 3.1.

Seja  $x' = (x_1, \dots, x_p)$  e  $y' = (y_1, \dots, y_p)$  temos que a Distância Euclidiana será dada por:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_p - y_p)^2}. \quad (3.17)$$

### 3.6. Métodos Hierárquicos

Os procedimentos utilizados para grupamentos nos métodos hierárquicos consistem em uma série de uniões ou divisões sucessivas de observações formando grupos, sendo que os resultados obtidos podem ser observados na forma de diagramas

bidimensionais, os dendogramas. Dentre os métodos hierárquicos podemos citar o método de *linkagem* simples, *linkagem* completa e método de *ward*.

### 3.7. Verificação da Heterogeneidade entre Grupos

A idéia da criação de grupamentos tem por objetivo a criação de classes cujos elementos possuem características homogêneas, sendo que estas classes são heterogêneas entre si. Para verificar estas a heterogeneidade entre os existe um teste que leva em consideração a suposição de normalidade, dado por:

**Hipóteses:**

**H<sub>0</sub>:** As médias entre os grupos são iguais para a  $i^{th}$  variável

**H<sub>1</sub>:** As médias entre os grupos não são iguais para a  $i^{th}$  variável

Assim, o teste é feito para todas as variáveis e espera-se na maioria das vezes que grupos diferentes possuam características diferentes entre si.

A estatística de teste é dada por  $F_i = \frac{QMEC}{QMDC}, i = 1, 2, \dots, p,$

$$QMEC = \frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2}{k - 1}, \text{ e } QMDC = \frac{\sum_{i=1}^K \sum_{j=1}^k (x_{ij} - \bar{x}_i)^2}{n - k}, \text{ sendo } k \text{ o número de grupos}$$

formados.

### 3.8. Identificando o Número de Grupos

Quando técnicas de análise de *cluster* são utilizadas, um dos maiores problemas é a identificação e escolha do número de grupos a serem formados.

Uma maneira para determinar o número de grupos é por métodos gráficos (os dendogramas), criados a partir da metodologia hierárquica, como já foi descrito anteriormente (seção 3.6). Assim, deve-se observar nos gráficos onde existe o maior “salto” e realizar um “corte” e tomar o número de grupos formados abaixo deste ponto. Além dos métodos gráficos existem outras maneiras numéricas de se encontrar o número de grupos, como por exemplo, a utilização do  $R^2$ , pseudo F e a estatística CCC.

## Capítulo 4

### 4.1. Considerações Finais

Até o momento foi realizado um estudo de uma maneira simples e coerente de realizar as estimações descritas no Capítulo 2 assim como entender melhor como se dá o processo de construção de índices e o que são algumas covariáveis utilizadas.

Utilizando as metodologias descritas no Capítulo 3, espera-se encontrar modelos que possibilitem a predição de diversos tipos de indicadores utilizados na construção civil. Propõe-se ainda identificar dentre as covariáveis estudadas, quais são as que realmente impactam nos indicadores e o quão importantes são para as estimações.

## Bibliografia

- [1] FUNDAÇÃO JOÃO PINHEIRO, *Diagnóstico nacional da indústria da construção*, Belo Horizonte, 1984.
- [2] FARAH, M. S. F. Diagnóstico tecnológico da indústria da construção civil: Caracterização geral do setor. In: *Tecnologia de Edificações*. São Paulo, Pini, 1992.
- [3] FREIRE, A. S., ; Indicadores de projeto para edifícios em alvenaria estrutural – *Tese de mestrado* - São Carlos: UFSCar, 2008
- [4] Martins, Guilherme Lopes; Sistema On-Line de Indicadores de Projeto em Edifícios de Alvenaria Estrutural – Benchmark – *Primeiro relatório de Iniciação Científica* – São Carlos: UFSCar, 2009
- [5] NETER, John; WASSERMAN, William; KUTNER, Michael H.. *Applied linear regression models*. Homewood: IRWIN, 1983.
- [6] CRIBARI NETO, F. ; FERRARI, S. L. P. ; CORDEIRO, G. M. . Improved Heteroscedasticity-Consistent Covariance Matrix Estimators, *Biometrika* v. 87, n. 4, p. 907- 918, 2000.
- [7] Anscombe, F. J. Graphs in Statistical Analysis, *American Statistician*, 27 (February 1973), 17-21.
- [8] MARDIA, Kantilal Varichand; KENT, J.T.; BIBBY, J.M.. *Multivariate analysis*. London: Academic Press (Probability and Mathematical Statistic. A Series of Monographs and Textbooks)
- [9] Software, Minitab.

## Referências Complementares

- [10] JOHNSON, Richard Arnold; WICHERN, Dean W.. *Applied multivariate statistical analysis*. Englewood Cliffs: Prentice-Hall.