

# Estudo estatístico Análise de Regressão

Anderson Queiroz

Rodrigo Formiga

Senai Cimatec. CCRoSA - Centro de Competência em Robótica e

Sistemas Autônomos

[anderson.vale@fbter.org.br](mailto:anderson.vale@fbter.org.br)

[rodrigo.farias@fbter.org.br](mailto:rodrigo.farias@fbter.org.br)

22 de setembro de 2020

# 1 Revisão literária

Serão apresentados na presente seção alguns pontos que, segundo a literatura, nortearão o estudo estatístico do trabalho em questão, abordando alguns pontos importantes para realização do planejamento e análise dos experimentos. Logo em seguida, serão aplicados os conceitos aqui apresentados.

## 1.1 Planejamento e análise de experimentos

Técnicas de planejamento e análises de experimentos são utilizáveis em empresa, processo de fabricação, desenvolvimento de um novo produto e até em fabricação de commodities. Geralmente a maioria destes processos possuem diversas variáveis controláveis, como temperatura, velocidade, dimensões entre outras. A variação destas variáveis, de modo isolado ou combinado, pode trazer melhorias ou prejuízos a qualidade do produto ou serviço. Com isso, faz-se necessário um estudo analítico de como cada variável pode influenciar o processo [MONTGOMERY, 2012].

O planejamento e análise de experimentos (*Design of Experiments*, **DOE**) é uma técnica de planejamento de experimentos, ou seja, como os experimentos devem proceder para garantir conclusões assertivas. A validade das conclusões obtidas após análise dos experimentos, são diretamente ligadas ao modo como os experimentos foram realizados, logo, o planejamento do experimento torna-se o ponto mais importante na hora de se analisar algum processo produtivo, bens ou serviços prestados. O planejamento corretamente executado também pode evitar desperdícios, além de isolar e determinar relações das variáveis [MONTGOMERY, 2012].

A seguir serão apresentados alguns pontos importantes para análise estatística realizada neste presente trabalho, como: Princípios do planejamento, que irão nortear como deve ser condizido o planejamento de experimentos, bem como etapas para realização dos experimentos e análise dos mesmos [TAHARA, 2008].

### 1.1.1 Princípios do planejamento de experimentos

O planejamento de experimentos possui três principais princípios básicos, que são: Replicação, aleatoriedade e blocagem [TAHARA, 2008].

#### **Replicação**

A replicação consiste na obtenção de mais de uma unidade experimental, ou seja, replicar o teste para cada ponto experimental e assim permite que obtenha-se uma estimativa mais precisa, diminuindo a influência de variações indesejadas ou inevitáveis [TAHARA, 2008].

## Aleatoriedade

Os métodos estatísticos requerem que as observações sejam variáveis aleatórias, de modo a garantir a distribuição igual dos fatores não esperados. Com isso garantem-se estimativas não tendenciosas dos efeitos e erros experimentais, bem como evitar influência sistemática de fatores não controláveis [TAHARA, 2008].

## Blocagem

A blocagem é uma técnica extremamente importante, utilizada com o objetivo de aumentar a precisão de um experimento. Este princípio pode ser aplicado em alguns casos onde deseja-se isolar algum fator conhecido, analisando de forma individual sua influência sobre o processo. A mudança de pessoas no processo experimental ou a mudança de lote de um produto pode ser visto a princípio como um fator não desejado de ser analisado, logo, o princípio de blocagem permite que o mesmo seja isolado, porém, não ignorado [TAHARA, 2008].

### 1.1.2 Etapas para o desenvolvimento de experimentos

Coleman e Montgomery (1993) propõem as seguintes etapas para o desenvolvimento de um planejamento de experimentos [COLEMAN, 1993]:

1. **Caracterização do problema:** A definição do problema que está sendo analisado é uma etapa essencial para se entender o estudo analítico. Para isso se faz necessário conhecimento sobre todo o processo para assim definir de forma clara o objetivo e relatar de forma específica o problema analisado.
2. **Escolha dos fatores de influência e níveis:** Para conduzir o experimento deve-se escolher os fatores variáveis, os intervalos sobre os quais esses fatores variarão e os níveis específicos. Para tanto é necessário conhecimento do processo, experiências práticas e teóricas para determinar tais fatores mais relevantes para análise experimental.
3. **Seleção das variáveis de resposta:** É também necessário ao planejar um experimento que se tenha clareza de qual variável-resposta será obtida como parâmetro para qualificar os experimentos. Variável de resposta que não atenda a necessidade do experimento pode levar a conclusões equivocadas.
4. **Determinação de um modelo de planejamento de experimento:** A escolha do planejamento envolve consideração sobre o tamanho da amostra (número de replicações) e determinar se há formação de blocos ou outras restrições de aleatorização.

5. **Condução do experimento:** A condução do experimento deve seguir todo o planejamento, evitando variações de ambiente, metodologia do experimento ou inserção de novas variações não previamente estabelecidas. O responsável pelo experimento e os equipamentos utilizados devem ser mantidos do início ao fim, exceto em casos onde estas variações sejam desejáveis.

6. **Análise dos dados:**

Após todas as demais etapas anteriores, os resultados podem ser obtidos com uso de ferramentas e pacotes estatísticos para auxiliar na visualização de gráficos e dados. A verificação da validade do modelo é também um ponto importante para analisar.

7. **Conclusões e recomendações:**

Após a obtenção dos resultados, o experimento deve apresentar conclusões práticas para proporcionar recomendações de ações em cima do processo analisado. As recomendações são frutos das etapas decorridas no planejamento do experimento, ou seja, planejamentos com falhas iram gerar ações equivocadas no processo.

## 1.2 Análise de regressão

Um método muito utilizado nos estudos estatísticos é a análise de regressão. Este método permite examinar a relação entre duas ou mais variáveis que se deseja-se estudar. O método permite por meio de modelos matemáticos avaliar quais as relações entre as variáveis, quais delas são importantes para o processo e quais apresentam pouca relevância [Action, sd].

As variáveis utilizadas podem ser divididas em duas: Variável Dependente, que é a variável que está sendo estudada, como uma saída do processo (a influência das demais variáveis será analisada através das variáveis dependentes); e Variável Independente, que são as entradas do estudo, são variáveis que supostamente causam impactos ou certa influência na variável dependente [Action, sd].

A análise de regressão tem por finalidade chegar a algumas conclusões e direcionamentos sobre o processo estudado, as finalidades deste estudo são [Action, sd]:

- Predição dos dados;
- Seleção de variáveis influenciáveis no processo;
- Estimação de parâmetros;
- Realizar inferências sobre os parâmetros, como: testes de hipóteses e intervalos de confiança.

### 1.2.1 Tipos de regressão

A análise de regressão é dividida em alguns tipos, tendo cada uma sua própria especificidade, logo, todo analista deve saber qual forma usar, variando sua escolha pelo tipo de dado e sua distribuição. Os tipos mais comuns são: Regressão Linear; Regressão Polinomial; Regressão de Poisson; Regressão de Ridge; Regressão Logística e Mínimos quadrados parciais (PLS) [MARIA, sd]. Para fins deste estudo será abordado de forma mais aprofundada a regressão linear.

#### Regressão linear

A regressão Linear é um modelo matemático que tem por objetivo observar a relação entre duas ou mais variáveis por meio de uma reta, e utilizar o resultado da função dessa reta para estimar valores e encontrar relações. Quando existe apenas uma variável independente e uma variável dependente, ela é chamada de **regressão linear simples**. Quando existe mais de uma variável independente, é chamado de **regressão linear múltipla**. Na equação de regressão linear múltipla, (equação 1), o 'Y' é a variável dependente, 'X' são as variáveis independentes,  $\beta_i$  são os coeficientes de regressão e  $\epsilon$  é o termo de erro [MARIA, sd].

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_i X_i + \epsilon \quad (1)$$

Como exemplo de uma análise de regressão linear no R (figura 1), será analisado os dados a seguir e entender as informações contidas no estudo.

Figura 1: Exemplo de regressão linear no R [MARIA, sd]

```
(Intercept)      Agriculture      Examination      Education      Catholic
66.9151817      -0.1721140      -0.2580082      -0.8709401      0.1041153
Infant.Mortality
1.0770481
> summary(model)

Call:
lm(formula = Fertility ~ ., data = swiss)

Residuals:
    Min       1Q   Median       3Q      Max
-15.2743  -5.2617   0.5032   4.1198  15.3213

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  66.91518   10.70604   6.250 1.91e-07 ***
Agriculture  -0.17211    0.07030  -2.448  0.01873 *
Examination  -0.25801    0.25388  -1.016  0.31546
Education    -0.87094    0.18303  -4.758 2.43e-05 ***
Catholic      0.10412    0.03526   2.953  0.00519 **
Infant.Mortality 1.07705    0.38172   2.822  0.00734 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.165 on 41 degrees of freedom
Multiple R-squared:  0.7067, Adjusted R-squared:  0.671
F-statistic: 19.76 on 5 and 41 DF, p-value: 5.594e-10
```

## **R múltiplo**

Esse dado serve para medir a relação linear entre as variáveis dependente e independentes, ou seja, mede o quanto elas estão correlacionadas. O valor desejável é próximo de 1 (100%) [MARIA, sd].

## **R-Quadrado**

O R-quadrado é uma medida estatística de quanto os dados estão próximos em linha de regressão. Ele também é conhecido como coeficiente de determinação múltipla para a regressão múltipla. Este valor varia de 0-1, onde o 0(zero) indica que o modelo não serve para explicar a variabilidade dos dados, enquanto 1(um) indica que o modelo consegue explicar toda a variabilidade dos dados de resposta ao redor da média[MARIA, sd].

## **Coeficientes**

São os valores que serão multiplicados pelas variáveis independentes para obter o valor esperado da variável dependente. Valores próximos de 1(um) indicam que a variável independente analisada interfere fortemente na variável de saída[MARIA, sd].

## **P-value**

Servem para avaliar a hipótese nula. Em valores inferiores a 0.05 é rejeitada a hipótese nula e a variável não deve ser descartada ou ignorada para o estudo. No caso da figura 1, a variável "Education"apresenta p-value muito pequeno e menor que 0.05, logo, esta variável influencia fortemente a variável de saída[MARIA, sd].

## Referências

- [Action, sd] Action, P. (s.d.). Análise de regressão. <http://www.portaction.com.br/analise-de-regressao>.
- [COLEMAN, 1993] COLEMAN, D. E.; MONTGOMERY, D. C. (1993). *A systematic approach to planning for a designed industrial experiment*. Technometrics.
- [MARIA, sd] MARIA, J. (s.d.). 15 tipos de regressão mais frequentes. <https://rpubs.com/JulhinhaM/395633>.
- [MONTGOMERY, 2012] MONTGOMERY, Donald C.; RUNGER, G. (2012). *Estatística Aplicada e Probabilidade Para Engenheiros*. LTC; 5ª ed.
- [TAHARA, 2008] TAHARA, S. (2008). Melhores práticas - planejamento de experimentos (doe). <http://www.portaldeconhecimentos.org.br/index.php/por/Conteudo/Planejamento-de-Experimentos-DOE>.