



Estatística para pesquisa

UM GUIA RÁPIDO E PRÁTICO

Marco Reis <marcoreis@me.com>

Robotics & Autonomous Systems Center, Senai Cimatec

Janeiro de 2022

Sistema FIEB



PELO FUTURO DA INOVAÇÃO

Introdução

A metodologia é a base de todo desenvolvimento de um projeto. Em projetos de pesquisa a Estatística deve ser a mola propulsora.

METODOLOGIA			
Robótica		DMAIC	Científica
Conceitual	ideação	Define	faça uma pergunta
Detalhamento	definição		faça um estudo do estado da arte
	planejamento		construa uma hipótese
Desenvolvimento	integração e testes	Improve	teste a hipótese com um experimento
Conclusão	análise		analise os dados e construa uma conclusão
	documentação	Control	comunique os resultados

Classe Mundial

Descubra o que os clientes querem, e forneça-lhes sempre o mesmo produto (bem ou serviço), sem nenhuma variação (no prazo de entrega acertado, na qualidade combinada, na quantidade contratada, e no preço estabelecido).



Six Sigma - 6σ

UM BREVE HISTÓRICO

- Criado pela Motorola na década de 80
- Adotado mundialmente por diversas empresas: Allied Signal, General Electric, Polaroid, Asea Brown Boveri, Kodak, Citicorp, Nokia, Sony, Toshiba, etc.
- Six Sigma Academy (SSA) foi criada para estimular a sua divulgação
- Chega ao Brasil na segunda metade da década de 90
- Interesse crescente no mundo e no Brasil

Six Sigma - 6σ

DEFINIÇÃO

Seis Sigmas é uma **metodologia** para melhoria de processos que faz com que se atinjam níveis de defeitos de 3,4 ppm para as características críticas de qualidade dos clientes.

Mikel Harry

Seis Sigmas é uma **filosofia** de negócios que visa a obtenção de produtos virtualmente livres de defeitos.

Jack Welch

Six Sigma - 6σ

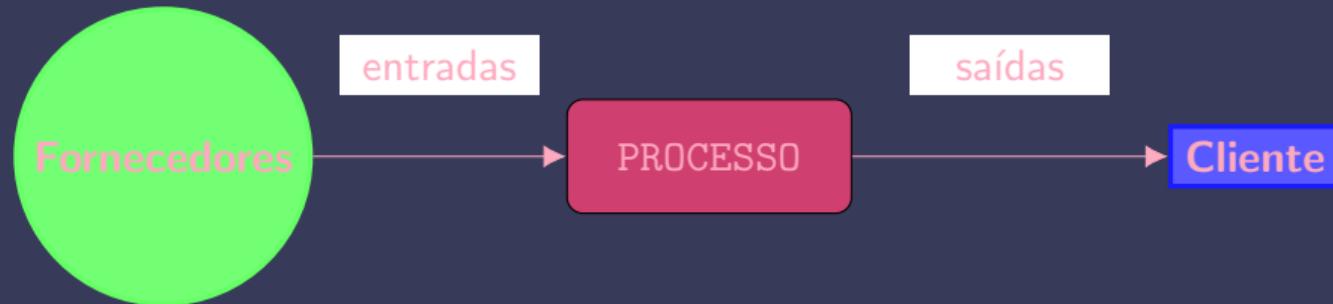
AINDA FALANDO DAS DEFINIÇÕES

- Seis Sigma é um mapa geral que ajuda a integrar as ferramentas que visam melhoria.
- Pode ser aplicado a todo tipo de processo: Finanças, RH, Vendas, Contabilidade, Manufatura, Administrativo (Transacional) etc.
- Todos usam a metodologia para melhoria, não só os especialistas.
- Ênfase na aplicação intensiva de ferramentas básicas e, quando necessárias, técnicas mais avançadas.
- O pensamento estatístico é utilizado largamente por todas pessoas da empresa.

A estatística

O pensamento estatístico é norteado pelos seguintes princípios:

- **todo trabalho** por nós executado é um conjunto de **processos** inter-relacionados
- todo processo está sujeito a **variação**, em maior ou menor quantidade
- a chave para a melhoria dos processos é atacar e **reduzir** sistematicamente a sua **variação**



Variações nas características críticas do produto causam insatisfação nos clientes.

Six Sigma - 6σ

CONCEITOS BÁSICOS

- Defeito é qualquer coisa que bloqueia ou inibe um processo de maior produtividade.
- Cliente (interno ou externo) são as pessoas ou empresas a quem se destinam as saídas do processo.
- Características críticas de qualidade (CTQ's) são características do produto essenciais ao cliente.
- Produtos são tanto bens como serviços.
- Diminuir a variação sem saber o que o cliente quer é pura perda de tempo.

O que é classe mundial?

Processo que gera 99% de produtos bons

- 20.000 itens perdidos por hora pelo correio;
- 15 minutos por dia de água não-potável;
- 5.000 cirurgias incorretas por semana;
- 2 aterrissagens incorretas por dia num grande aeroporto;
- 200.000 receitas médicas incorretas por ano;
- 7 horas de falta de energia elétrica por mês.

Um processo Seis Sigma

- 7 itens perdidos por hora pelo correio;
- 1 minuto de água não potável a cada 7 meses;
- 2 cirurgias incorretas por semana;
- 1 aterrissagem incorreta a cada 5 anos num grande aeroporto;
- 68 receitas médicas incorretas por ano;
- 1 hora de falta de energia elétrica a cada 34 anos.

Por que adotar o 6σ?

Para melhorar o negócio da empresa

- Resolução de Problemas
- Tomada de Decisão com base em Fatos

Para aumentar a eficácia dos investimentos

- Diminuir custos de falhas, defeitos, etc.
- Aumentar produtividade
- Aumentar valor da empresa no mercado

Para melhorar o relacionamento com os clientes

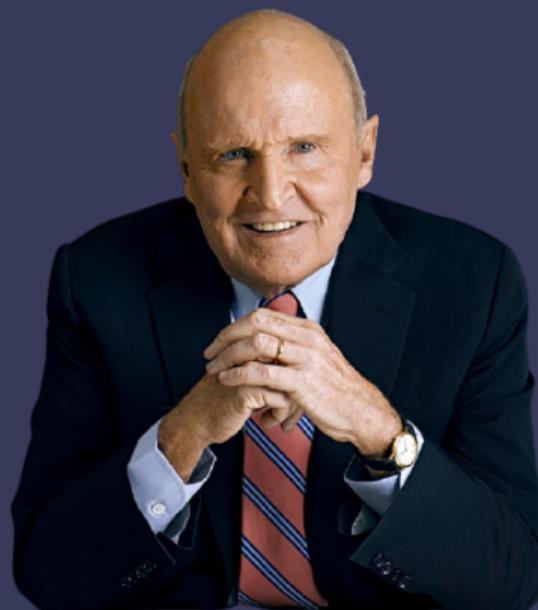
- Aumentar as vendas
- Vencer a concorrência

6σ

O 6σ de Jack Welch

"O Seis Sigma mudou o DNA da General Electric - é agora a forma que trabalhamos - em tudo o que fazemos e em todo produto que desenvolvemos."

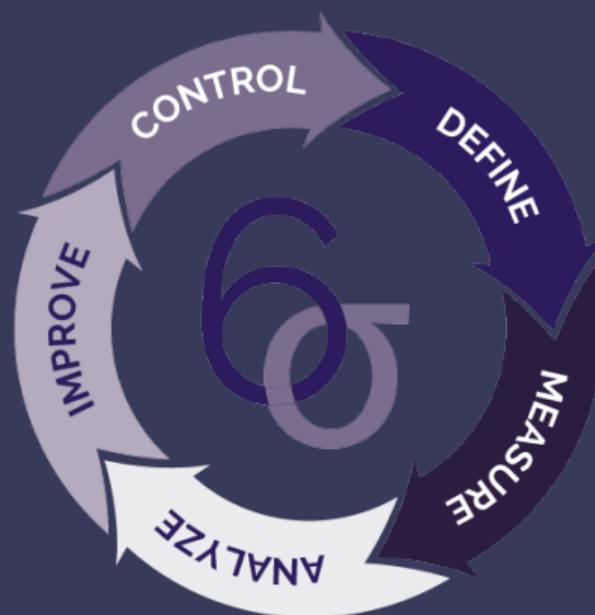
Jack Welch foi um executivo americano, autor de vários livros. Ficou 20 anos na liderança da General Electric, durante este tempo tirou a empresa de uma grande burocracia e aplicou diversas inovações gerenciais.
[Jack Welch – Wikipédia, a encyclopédia libre]



Voltando para a metodologia

As fases da metodologia DMAIC
[Zinkgraf e Snee 1999]

1. O que deve ser feito?
2. Como estamos atualmente?
3. O que é importante?
4. Como melhorar?
5. Como manter a melhoria?



D M A I C

A fase DEFINIÇÃO

Estabelecer os CTQ's dos clientes.

- Fatores que afetam a percepção que o cliente tem do produto.
- Requisitos específicos/técnicos
- Requisitos legais

Escutar a voz do cliente. A voz do cliente é algo extremamente importante para o desenvolvimento de um determinado produto/serviço. É a tradução das necessidades, expectativas e exigências dos clientes. Podemos definir clientes com internos ou externos.

Definir o projeto. Definir o escopo, a equipe; assim como identificar os processos-chave a serem relacionados ao desenvolvimento.

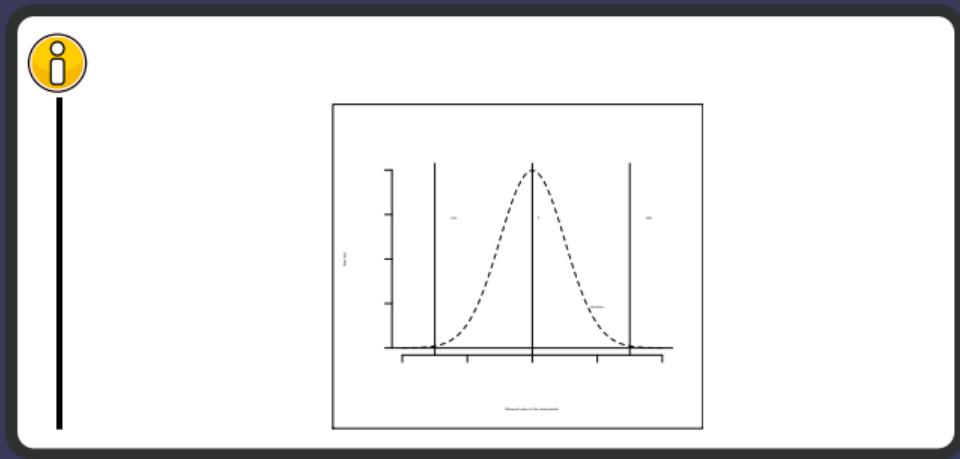
A fase MEDIÇÃO

Definir o processo: entradas, etapas e saídas

Coleta de dados do desempenho do processo

- avaliação do sistema de medição
- planejamento e avaliação dos dados

Determinação da capacidade do processo



- Redução da quantidade de variáveis candidatas:
 $y = f(x_1, x_2, \dots, x_k)$
- Análise do modo e efeito das falhas - FMEA
- Uso de ferramentas básicas e avançadas:
 - diagrama de Pareto
 - diagrama de dispersão
 - gráfico linear
 - histograma
 - *box-plot*
 - gráfico de controle
 - análise de regressão
 - estudo de correlação
 - análise de variância
 - testes de hipóteses

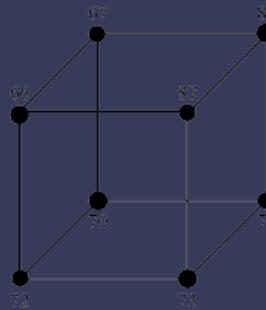
A fase

ANÁLISE

A fase MELHORIA

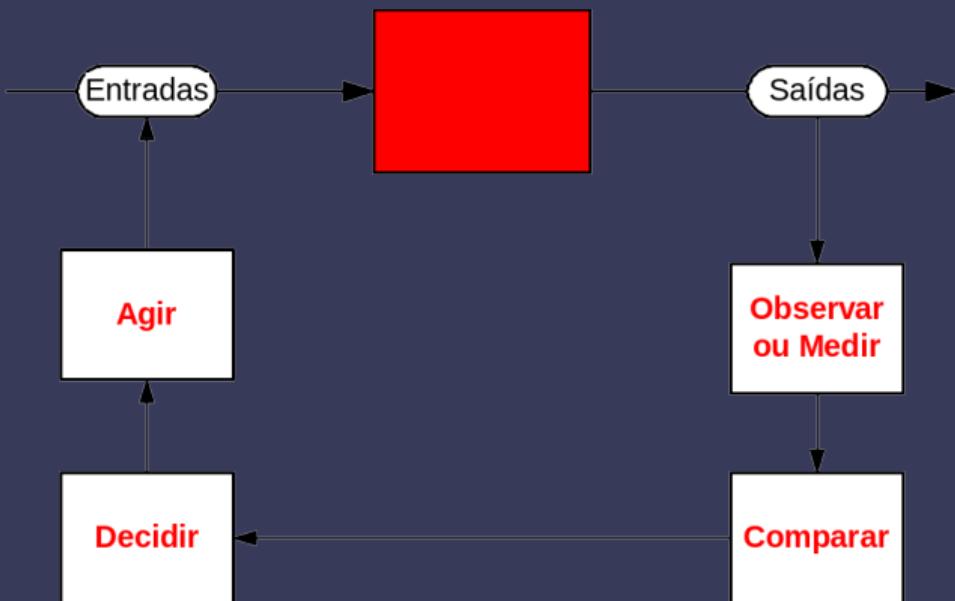
Experimentação científica

- Planejar execução da experiência
- Identificar variáveis importantes
- Avaliar interações
- Reduzir influências externas
- Otimizar os resultados
- Obter a melhor combinação

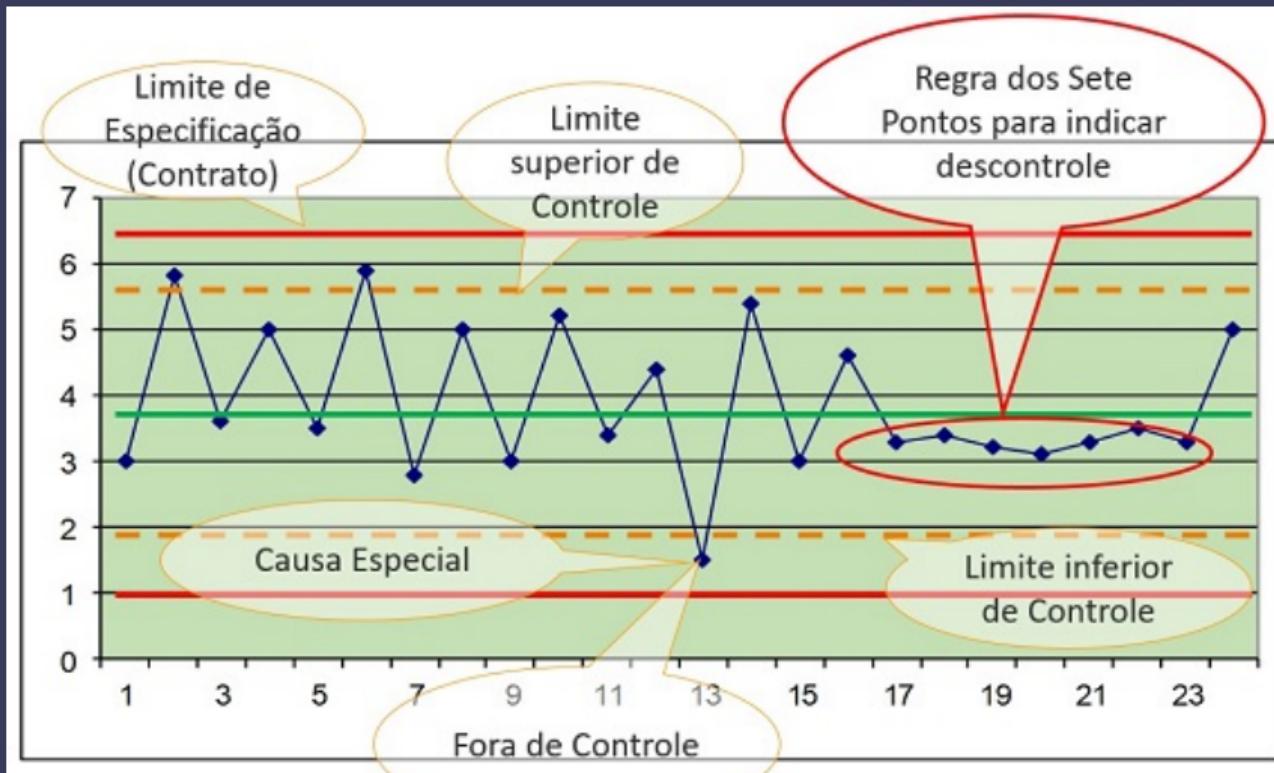


A fase CONTROLE

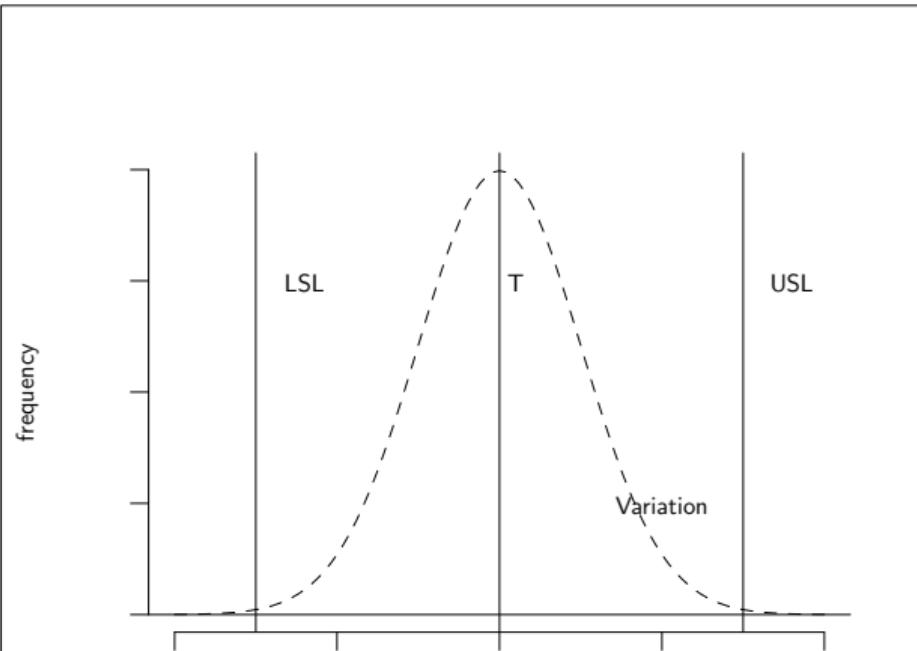
- Identificar as variáveis a serem controladas
- Padronizar processo



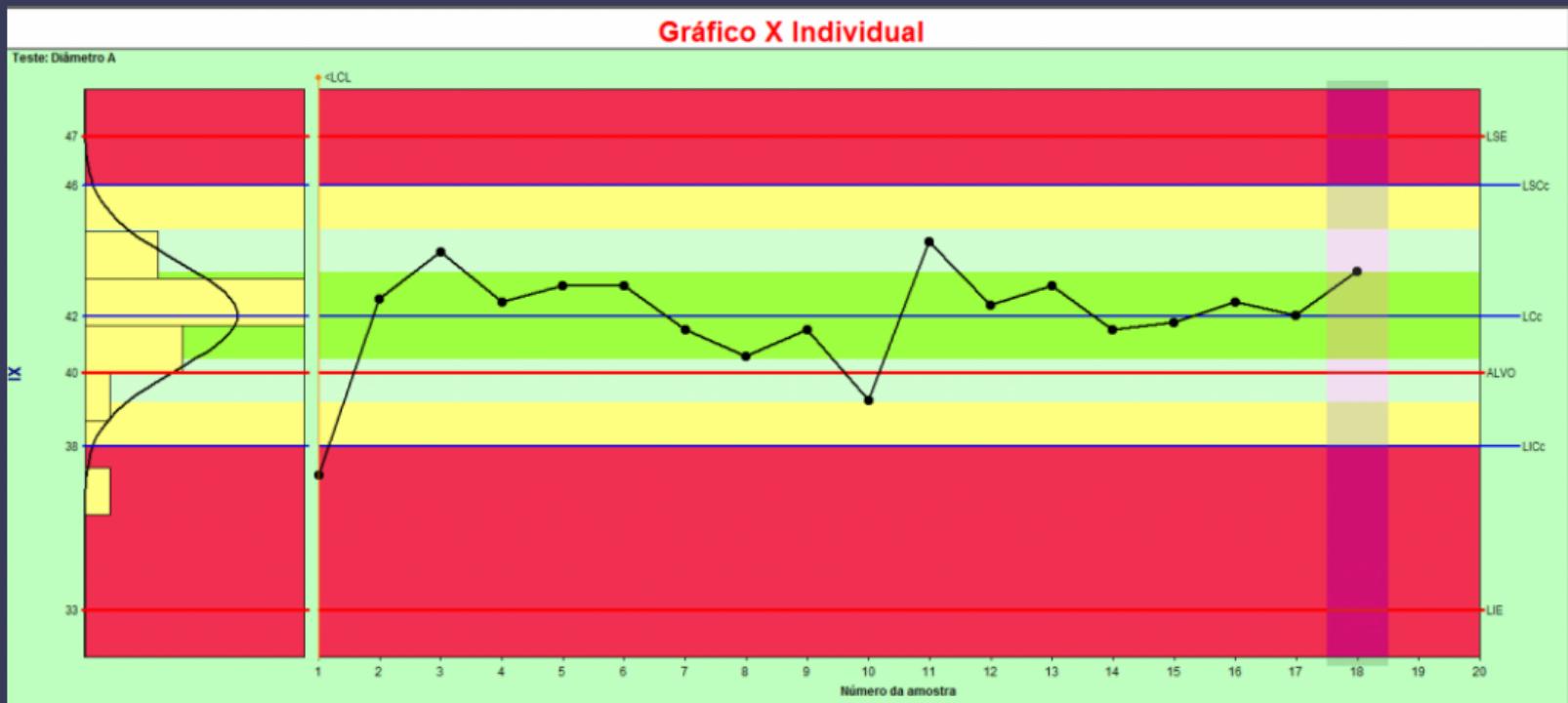
A universalidade da variação



A distribuição normal



Variação e o conceito da normalização



Aquele que ama a prática
sem teoria é como o
marinheiro que embarca no
navio sem leme e sem
bússola, e nunca sabe para
onde pode ser lançado.

Leonardo Da Vinci



O R Commander

É um software livre desenvolvido por John Fox; é uma interface gráfica do **R**. Pode ser considerado um recurso em contra-posição aos softwares comerciais, tais como o **SPPS** e **Minitab**.

O R Commander já está configurado para exportar dados gráficos para o Latex e Html.



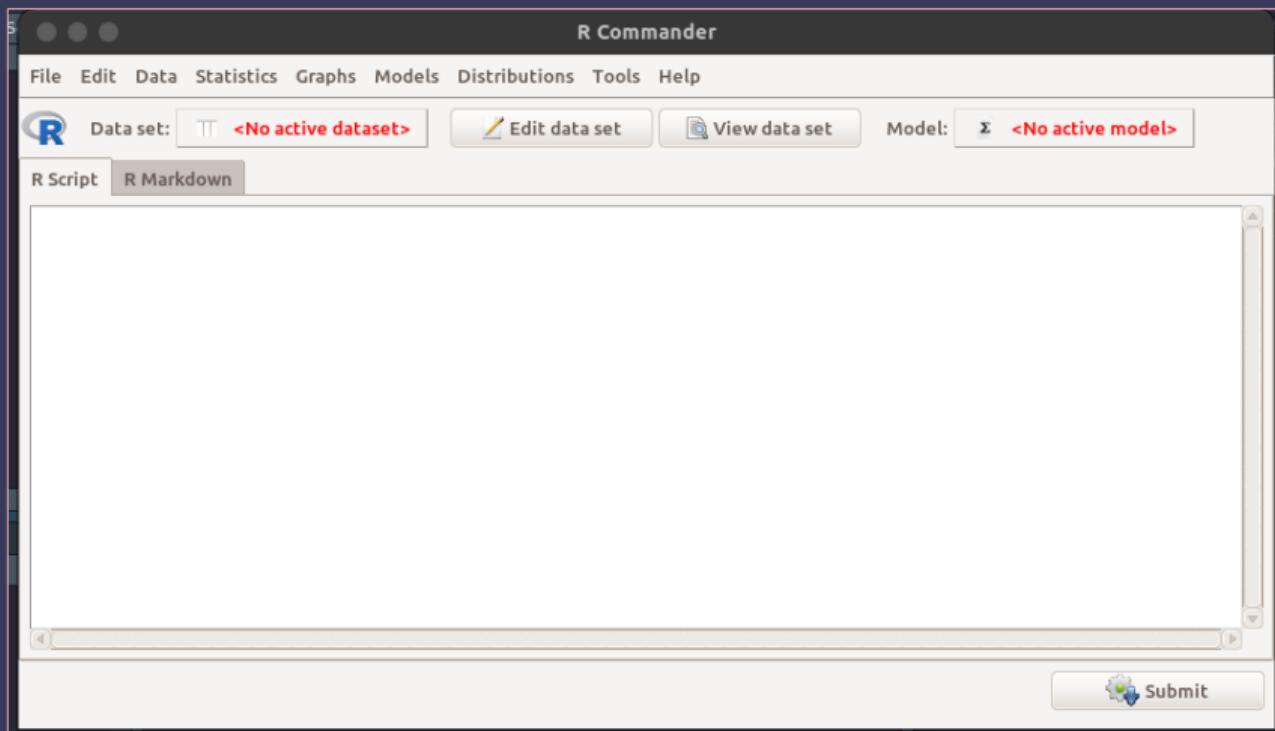
John Fox - Home Page

"Eu sou um sociólogo e estatístico social, então não tenho uma carreira na ciência, como esse termo é geralmente entendido. No entanto, eu estava interessado em ciências quando criança: frequentei uma escola de ciências em Nova York (Brooklyn Tech) e, quando comecei a universidade em 1964, no City College de Nova York, comecei em engenharia. Posteriormente, me formei em filosofia e psicologia, antes de terminar em sociologia – se não tivesse me formado em 1968, provavelmente teria feito outra coisa."



Uma entrevista com John Fox

Rcmdr na prática



O início da linguagem R

- não é um *app*, é uma linguagem de programação
- teve origem no Bell Laboratories na década de 70
- é uma evolução da linguagem **S**
- focado na pesquisa aplicada
- grande impulso na década de 90
- orientado a objetivo
- open source
- crescimento dos pacotes apartir de 2000
- Versão atual 3.6.3



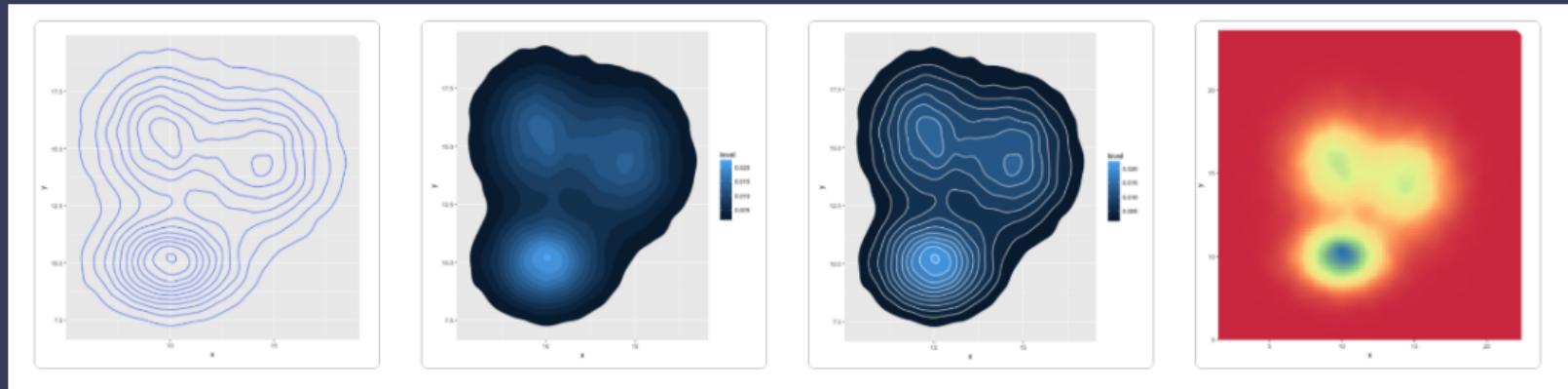
ggplot2

UM PACOTE DE DESTAQUE

O pacote é dedicado a visualização de dados.

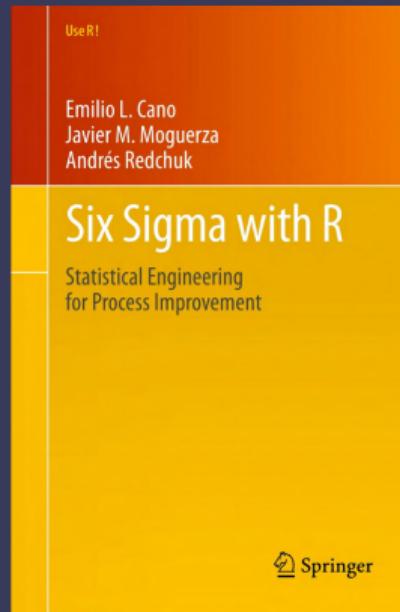
Ele pode melhorar muito a qualidade e a estética dos gráficos e o que torna a pesquisa desenvolvida muito mais eficiente e visual.

Neste [link](#) pode se encontrar vários modelos de gráficos prontos para serem utilizados.



R

COLOCANDO EM PRÁTICA



Passos iniciais



Para os primeiros passos com o R iremos utilizar um banco de dados com informações sobre a planta **Iris**[Iris (plant) - Wikipedia].

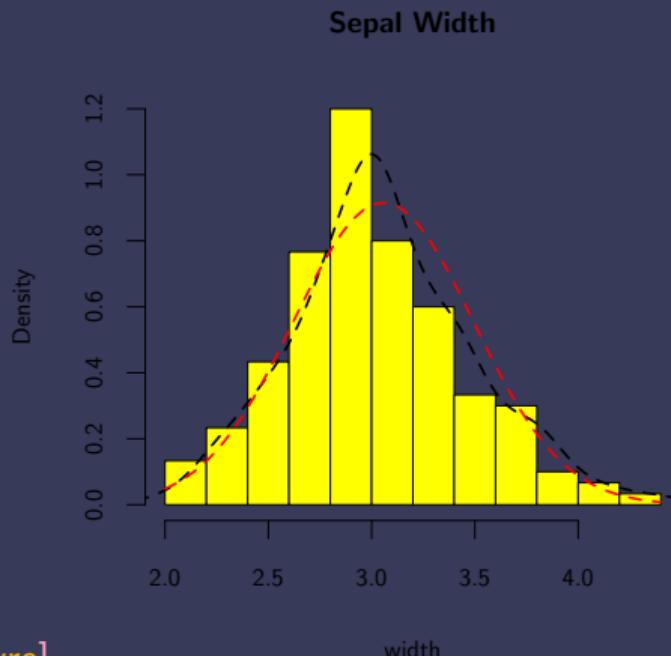
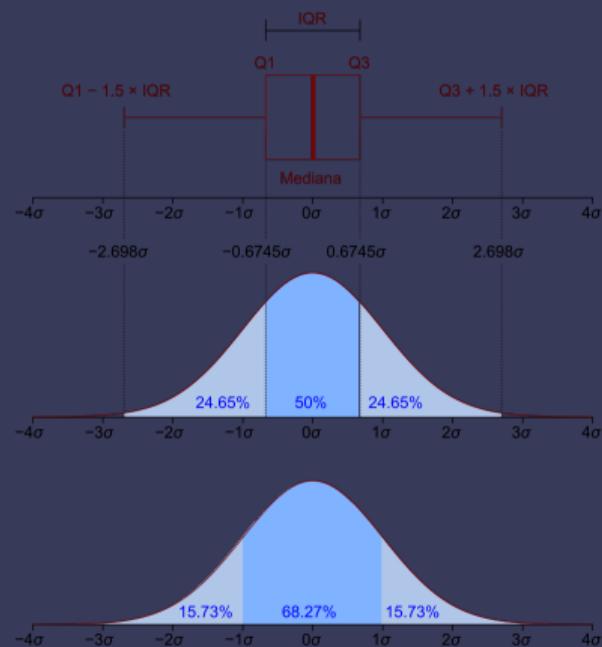


Os dados podem ser obtidos deste link.

Data Package Iris

```
# Inserir dados "CSV" em um data frame.  
> data.iris <- read.csv("iris.csv", header = TRUE, sep = ",")  
# Realizar teste de normalidade  
> shapiro.test(data.iris$sepal_width)  
# Informações básicas dos dados  
> summary(data.iris)  
# Gerar histograma  
> hist(data.iris$sepal_width, freq = FALSE, col = "yellow")  
# Gerar função densidade dos dados  
> lines(density(data.iris$sepal_width), lty = 2, lwd = 2)  
# Gerar função probabilidade dos dados  
> curve(dnorm(x, 3.057, 0.4358663), add = TRUE, lty = 2, lwd = 2)  
# Fazer um "split" do data frame.  
> data.iris.species <- split(data.iris, f=factor(data.iris$species))
```

A função Densidade + Diagrama de Caixa + Distribuição Normal



[Função densidade – Wikipédia, a encyclopédia livre]

Exportando figuras para o Latex

```
# Procedimento para gerar figura Tikz para o Latex.  
> tikz('funcao1.tex', width = 3.5, height = 3.5)  
> hist(data.iris$sepal_width, main = "Sepal Width", xlab = "width", freq =  
      FALSE, col = "yellow")  
> lines(density(data.iris$sepal_width), lty = 2, lwd = 2)  
> curve(dnorm(x, 3.057, 0.4358663), add = TRUE, col = "red", lty = 2, lwd = 2)  
> dev.off()
```

```
% Gerando figuras no Latex oriundas do R  
\begin{tikzpicture}[thick, scale=1, every node/.style={scale=0.75}]  
  \node [at=(current page.center)]  
  {  
    \input{./Source/r-graphics/funcao1.tex}  
  };  
\end{tikzpicture}
```

Introdução ao RStudio



O desenvolvimento será baseado no pacote do R:
SixSigma.

The screenshot shows the RStudio interface with several windows open:

- Code Editor (Top Left):** Shows R code for generating random numbers and creating a pie chart. A red circle labeled "1" highlights the first few lines of code.
- Environment Window (Top Right):** Displays the global environment with various objects like `cpus`, `data.iris`, and `my.list`. A red circle labeled "2" highlights the `data.iris` object.
- Console Window (Bottom Left):** Shows the R console output, including the command to load the iris dataset and the resulting pie chart error messages. A red circle labeled "3" highlights the error messages.
- Packages Window (Bottom Right):** Shows the User Library with a list of installed packages. A red circle labeled "4" highlights the `grid` package.

```

447 ## code chunk number 61: cld8_0nw:1305-1307
448 #####
449 factorial(5)
450 choose(5, 2)
451
452
453 #####
454 ## code chunk number 62: cld8_0nw:1322-1323
455 #####
456 rnorm(5, 5.2, 0.5)
457
458
459 #####
460 ## code chunk number 63: cld8_0nw:1338-1331
461 #####
462 rnorm(5, 5.2, 0.5)
463
464
465 #####
466 ## code chunk number 64: cld8_0nw:1340-1342
467 #####
468 qnorm(0.9, 5.2, 0.5)
469 dnorm(5.6, 5.2, 0.5)
470
471

```

```

1:1 [Top Level] : R Script
Console Terminal Jobs
R 3.6.3 ~/R/ ↵
140 149 6.9 3.1 5.4 2.1 virginica
141 140 6.7 3.1 5.6 2.4 virginica
142 141 6.9 3.1 5.1 2.3 virginica
143 142 5.8 2.7 5.1 1.9 virginica
144 143 6.8 3.2 5.9 2.3 virginica
145 144 6.7 3.3 5.7 2.5 virginica
146 145 6.7 3.0 5.2 2.3 virginica
147 146 6.3 2.5 5.0 1.9 virginica
148 147 6.5 3.0 5.2 2.0 virginica
149 148 6.2 3.4 5.4 2.3 virginica
150 149 5.9 3.0 5.1 1.8 virginica

> pie(data.iris$species$petal_width, labels = data.iris$species$species)
Error in pie(data.iris$species$petal_width, labels = data.iris$species$species) :
  'x' values must be positive.
> pie(data.iris$species$versicolor)
Error in pie(data.iris$species$versicolor) : 'x' values must be positive.
> pie(data.iris$petal_length, labels = data.iris$species)
> pie(data.iris$species, labels = data.iris$petal_length)
Error in pie(data.iris$species, labels = data.iris$petal_length) :
  'x' values must be positive.
> pie(data.iris$species, labels = data.iris$petal_length)
Error in pie(data.iris$species, labels = data.iris$petal_length) :
  'x' values must be positive.
> pie(data.iris$setosa$petal_width, labels = data.iris$species)
> View(data.iris$setosa)
> View(data.iris$setosa)
> View(data.iris)
> View(data.iris)
> pie(data.iris$setosa$petal_width, labels = data.iris$species)
>

```

Name	Description	Version
grid	Another Plot Package: Bagplots, 'Iconplots', 'Summaryplots', Slider Functions and Others	1.3.5
arm	Data Analysis Using Regression and Multilevel/Hierarchical Models	1.12-2
askpass	Safe Password Entry for R, Git, and SSH	1.1
backports	Reimplementations of Functions Introduced Since R 3.0.0	1.4.1
base64enc	Tools for base64 encoding	0.1-3
bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.0.4
bit64	A S3 Class for Vectors of 64bit Integers	4.0.5
brew	Templating Framework for Report Generation	1.0-6
brio	Basic R Input Output	1.1.3
broom	Convert Statistical Objects into Tidy Tibbles	0.7.11
cachem	Cache R Objects with Automatic Pruning	1.0.6
callr	Call R from R	3.7.0
car	Companion to Applied Regression	3.0-12
carData	Companion to Applied Regression Data Sets	3.0-5
caret	Classification and Regression Training	6.0-90
cellranger	Translate Spreadsheet Cell Ranges to Rows and Columns	1.1.0
checkmate	Fast and Versatile Argument Checks	2.0.0
cli	Helpers for Developing Command Line Interfaces	3.1.0
clipr	Read and Write from the System Clipboard	0.7.1

Um pouco de inferência

```
# Criar uma amostra de dados normal apartir da media e desvio padrao
> rnorm(5, 5.2, 0.5)
#
# Encontrar a probabilidade de encontrar um valor menor que 5.6 da amostra
# acima
> pnorm(5.6, 5.2, 0.5)
#
# Encontrar o primeiro quartil da amostra
> qnorm(0.25, 5.2, 0.5)
#
# Encontrar a densidade de probabilidade do valor 5.6 da amostra
> dnorm(5.6, 5.2, 0.5)
```

Gerando um gráfico de pizza

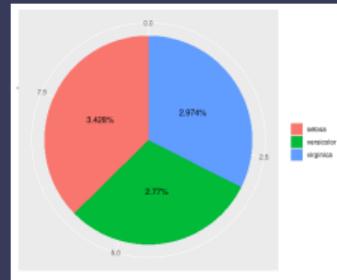
Utilizando o dataframe da flor **Iris**, deve-se organizar o *data set* para q o gráfico seja gerado.

Um dos melhores pacotes para manipular dados é o **dplyr**...

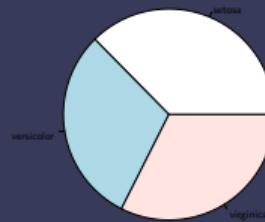
```
> new.data.iris <- data.iris %>%
+   group_by(species) %>%
+   summarise(across(starts_with("Sepal"), list(mean = mean, sd = sd)))
> new.data.iris
# A tibble: 3 x 5
species      sepal_length_mean sepal_length_sd sepal_width_mean sepal_width_sd
<fct>            <dbl>          <dbl>           <dbl>          <dbl>
1 setosa         5.01          0.352           3.43          0.379
2 versicolor     5.94          0.516           2.77          0.314
3 virginica      6.59          0.636           2.97          0.322
```

Gerando um gráfico de pizza

```
> ggplot(new.data.iris, aes(x="", y=
  sepal_width_mean, fill=species)) +
  geom_bar(stat="identity", width=1)
  +
  coord_polar("y", start=0) +
  geom_text(aes(label = paste0(sepal_
    width_mean, "%")), position =
    position_stack(vjust=0.5)) +
  labs(x = NULL, y = NULL, fill =
    NULL)
```



```
> pie(new.data.iris$sepal_width_mean,
  new.data.iris$species)
```



References (1)

[Função densidade – Wikipédia, a enclopédia livre] FUNÇÃO densidade – Wikipédia, a enclopédia livre. https://pt.wikipedia.org/wiki/Fun%C3%A7%C3%A3o_densidade#cite_note-3. (Accessed on 01/12/2022).

[Iris (plant) - Wikipedia] IRIS (plant) - Wikipedia. [https://en.wikipedia.org/wiki/Iris_\(plant\)](https://en.wikipedia.org/wiki/Iris_(plant)). (Accessed on 01/11/2022).

[Jack Welch – Wikipédia, a enclopédia livre] JACK Welch – Wikipédia, a enclopédia livre. https://pt.wikipedia.org/wiki/Jack_Welch. (Accessed on 01/10/2022).

References (2)

[Zinkgraf e Snee 1999] ZINKGRAF, A.; SNEE, R. Institutionalizing six sigma in large corporations: a leadership roadmap. In: **Quality and Productivity Research Conference**. [S.l.: s.n.], 1999. v. 20.



Questions?

marcoreis@fieb.org.br