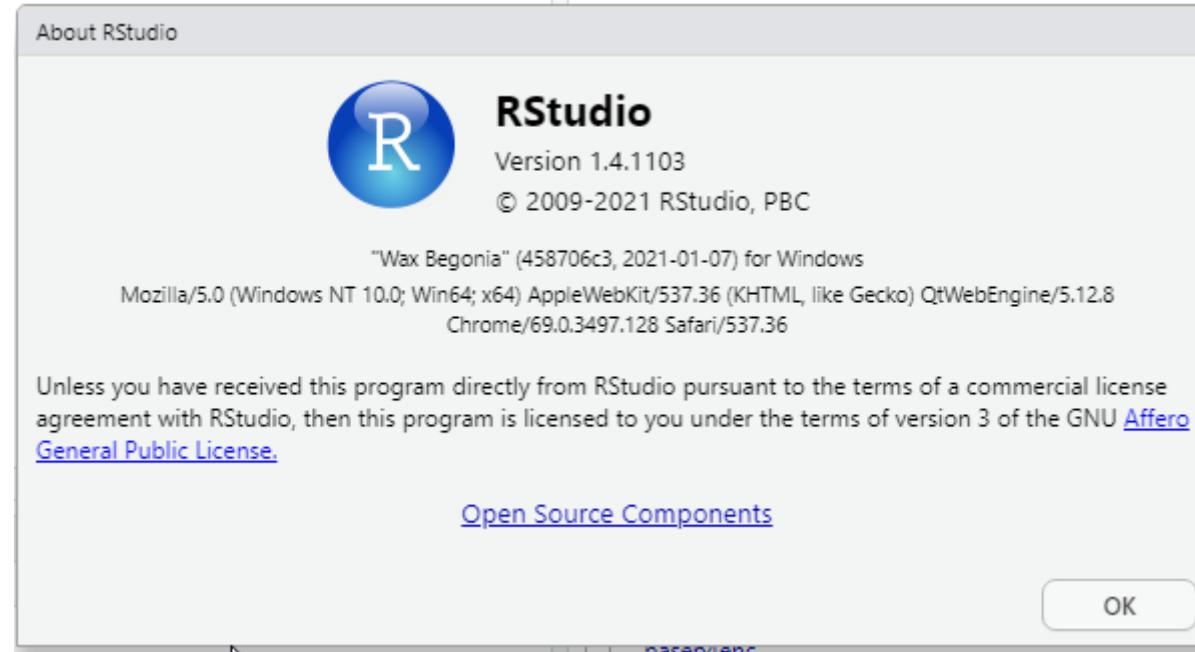


APOSTILA RStudio V1.4 com R V4.0.3



R version 4.0.3 (2020-10-10) -- "Bunny-Wunnies Freak Out"
Copyright (c) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

Treinamento criado por: Eduardo Santos
Aplicação em R criado por: Mário P. Rodrigues Jr

Índice Geral

Cap.	Assunto	Pág.
	Objetivo / Público alvo	3
	Introdução ao RStudio	4
1	Instalando R e RStudio (Nível: Básico)	5
2	Apresentando o RStudio (Nível: Básico)	16
3	Entrada e manipulação de dados (Nível: Básico)	22
4	Análise gráfica (Nível: Básico)	29
5	Estudo R&R de variáveis (Nível: Intermediário)	48
6	Nível Sigma (Atributos e Variáveis) (Nível: Intermediário)	56
7	Ferramentas do planejamento da qualidade (Nível: Intermediário)	59
8	Teste de hipóteses para dados variáveis (Nível: Avançado)	73
9	Teste de hipóteses para proporções (atributos) (Nível: Avançado)	85
10	Regressão (Nível: Avançado)	90
11	Análise de variância (Nível: Avançado)	101
12	Experimentos fatoriais completos (Nível: Mestre)	108
13	Experimentos fatoriais fracionados (Nível: Mestre)	128
14	Otimização de respostas (Nível: Mestre)	144
15	Planejamento de superfície de resposta (Nível: Mestre)	156
16	Cartas de controle (Variáveis) (Nível: Avançado)	173
17	Cartas de controle (Atributos) (Nível: Avançado)	183
18	Análise da tendência (Nível: Intermediário)	201
19	Glossário linguagem Six Sigma (Nível: Básico)	206
	Bibliografia	227

OBJETIVO

Fornecer material de suporte para a análise estatística dos dados

PÚBLICO ALVO

Líderes de projetos (Black Belts e Green Belts) e equipes de projetos Six Sigma, PDCA, TPM, etc.

Histórico

R é uma linguagem de programação multi-paradigma orientada a objetos, programação funcional, dinâmica, fracamente tipada, voltada à manipulação, análise e visualização de dados. Foi criado originalmente por Ross Ihaka e por Robert Gentleman no departamento de Estatística da Universidade de Auckland, Nova Zelândia. Atualmente é mantido por uma comunidade de colaboradores voluntários que contribuem com código fonte da linguagem e com a expansão de funcionalidades por bibliotecas.

O nome R provém em parte das iniciais dos criadores (Ross e Robert) e também de um jogo figurado com a linguagem S (da Bell Laboratories, antiga AT&T).

R é um ambiente e uma linguagem de programação similar ao S, contudo, é uma implementação distinta do S. Muitos códigos escritos para o S podem ser executados inalterados no R e vice-versa. A implementação comercial de S é S-PLUS.

O código fonte do R está disponível sob a licença GNU, GPL e as versões binárias pré-compiladas são fornecidas para Windows, Macintosh, e muitos sistemas operacionais Unix/Linux.

Principais ferramentas

- Estatística básica e avançada
- Testes de hipóteses e análise de variância (ANOVA)
- Análise de séries temporais
- Gráficos novos e otimizados com qualidade para apresentação
- Correlação e regressão
- Simulações e distribuições probabilísticas
- Importação, exportação e manipulação flexível de dados
- Controle estatístico de processos (CEP)
- Projeto de experimentos (DOE)
- Análise de confiabilidade
- Análise multivariada
- Estatística não-paramétrica
- Tamanho amostral e cálculo do poder do teste, etc...



1

Instalando o R e o RStudio

Nível: Básico

Instalação do interpretador R

Com seu navegador vá até a página r-project.org e clique em **Download R**



[Home]

[Download](#)

[CRAN](#)

[R Project](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

Escolha um mirror mais próximo de você:

Brazil

<https://nbcgib.uesc.br/mirrors/cran/>
<https://cran-r.c3sl.ufpr.br/>
<https://cran.fiocruz.br/>
<https://vps.fmvz.usp.br/CRAN/>
<https://brieger.esalq.usp.br/CRAN/>

Escolha o Sistema operacional:

Download and Install R

Precompiled binary distributions of the base system and contributed packages. Windows and Mac users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Clique em base

Subdirectories:

[base](#)

[contrib](#)

[old_contrib](#)

[Tools](#)

Binaries for base distribution. This is what you want to [install R for the first time](#).

Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.

Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).

Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

Faça o Download

[Download R 4.0.3 for Windows](#) (85 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

Instale normalmente como qualquer outro software.

Caso o seu computador bloqueie instalações por não ser administrador, use a versão portable.

<https://sourceforge.net/projects/rportable/files/latest/download>

Você terá que “instalar” da mesma forma, mas esse não pede privilégios de administrador.

Caso prefira, “instale” e um pendrive.

Instalação da IDE RStudio

Com seu navegador vá até a página rstudio.com e clique em **Download**

The screenshot shows the top navigation bar of the RStudio website. It includes the RStudio logo, a search icon, and links for DOWNLOAD, SUPPORT, DOCS, and COMMUNITY.

Escolha a versão FREE:

RStudio Desktop

Open Source License

Free

**Ou vá direto
No link abaixo:**

[DOWNLOAD](#)

[Learn more](#)

<https://rstudio.com/products/rstudio/download/#download>

**Se não estiver no
Windows, role a tela
Para baixo e veja:**

OS	Download
Windows 10/8/7	RStudio-1.4.1103.exe
macOS 10.13+	RStudio-1.4.1103.dmg
Ubuntu 16	rstudio-1.4.1103-amd64.deb
Ubuntu 18/Debian 10	rstudio-1.4.1103-amd64.deb
Fedora 19/Red Hat 7	rstudio-1.4.1103-x86_64.rpm
Fedora 28/Red Hat 8	rstudio-1.4.1103-x86_64.rpm
Debian 9	rstudio-1.4.1103-amd64.deb
OpenSUSE 15	rstudio-1.4.1103-x86_64.rpm

Instale normalmente como qualquer outro software.

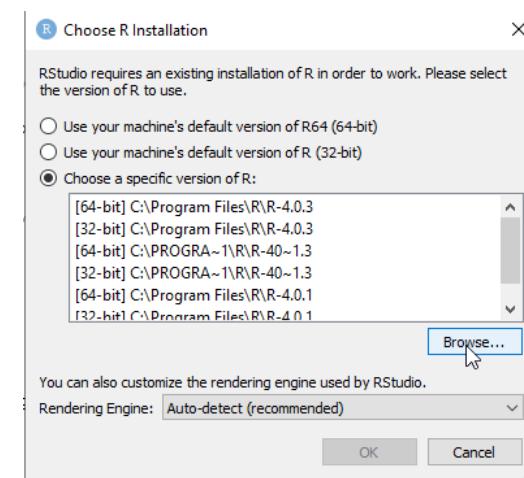
Caso o seu computador bloqueie instalações por não ser administrador, faça o seguinte:

Instale o rstudio no seu computador pessoal
Copie a pasta de instalação para o pendrive.



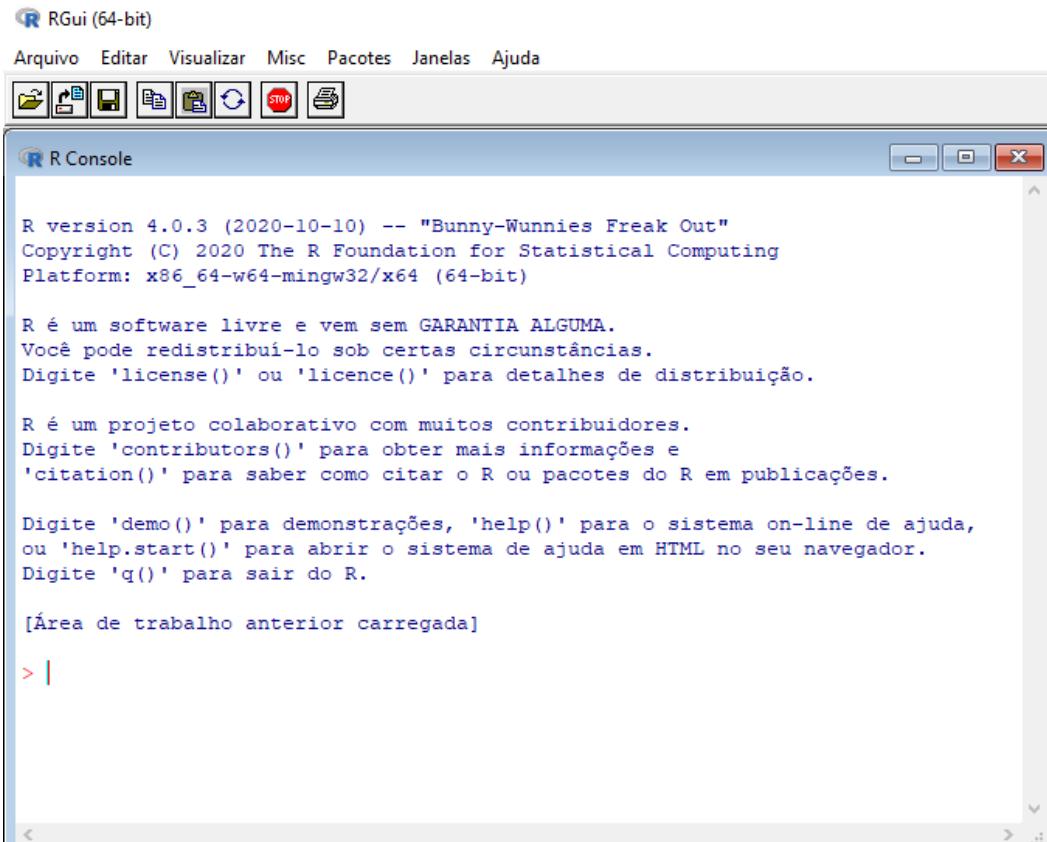
Se eu Windows for em português, Program Files poderá ser Arquivos de Programas.

Caso esse seja seu caso, ao abrir o Rstudio, ele dirá que não achou a instalação do R. Ness caso, clique em Browse e aponte para a sua pasta do R



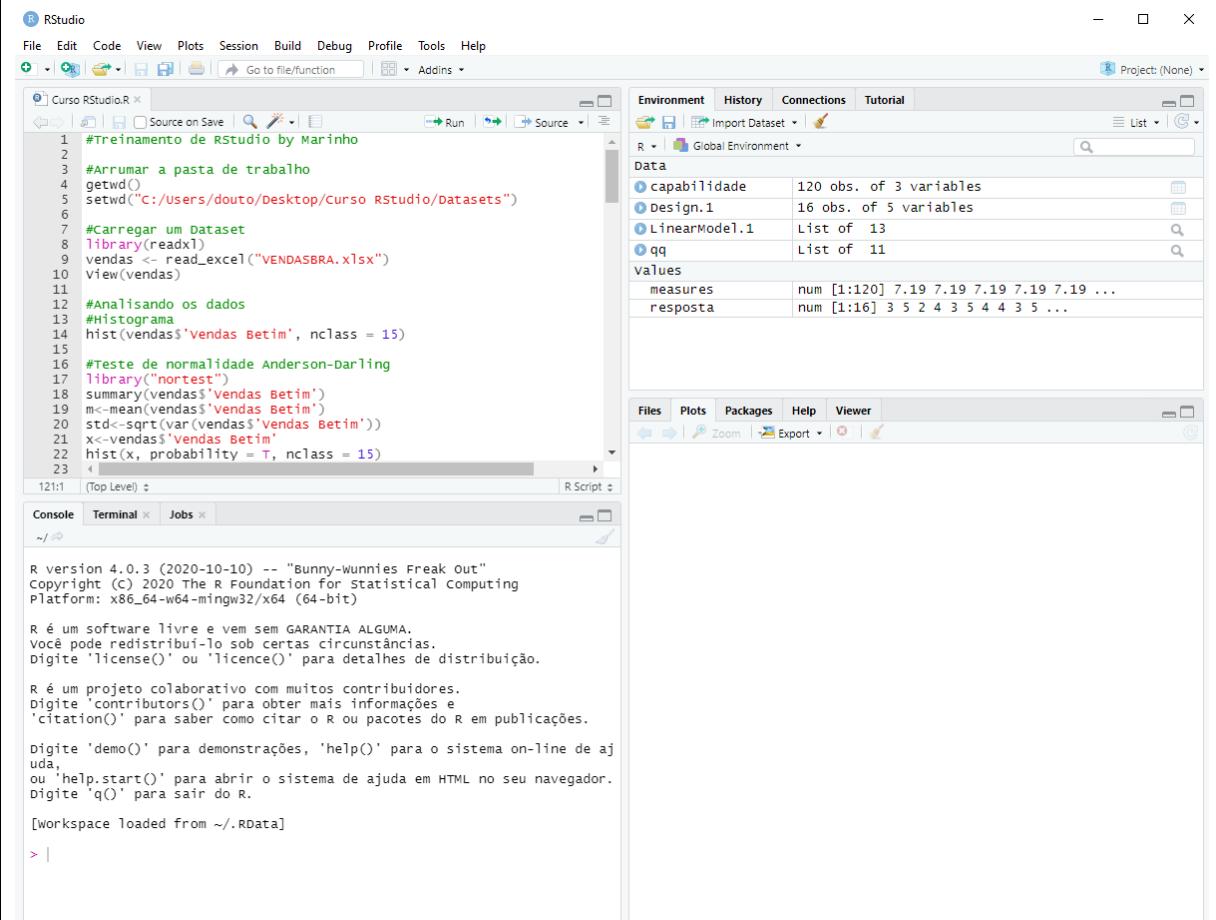
Abrindo RGUI e RStudio

Junto com a instalação do R temos o RGUI (R Graphical User Interface).



Ele nada mais é do que um prompt um pouquinho mais amigável.

Mas como instalamos o RStudio, podemos ter um ambiente bem mais completo:



Agora, vamos conhecer esse ambiente!



Apresentando o RStudio

Nível: Básico

Apresentando o RStudio

Quando o RStudio for inicializado, aparecerá uma tela idêntica a detalhada ao lado.

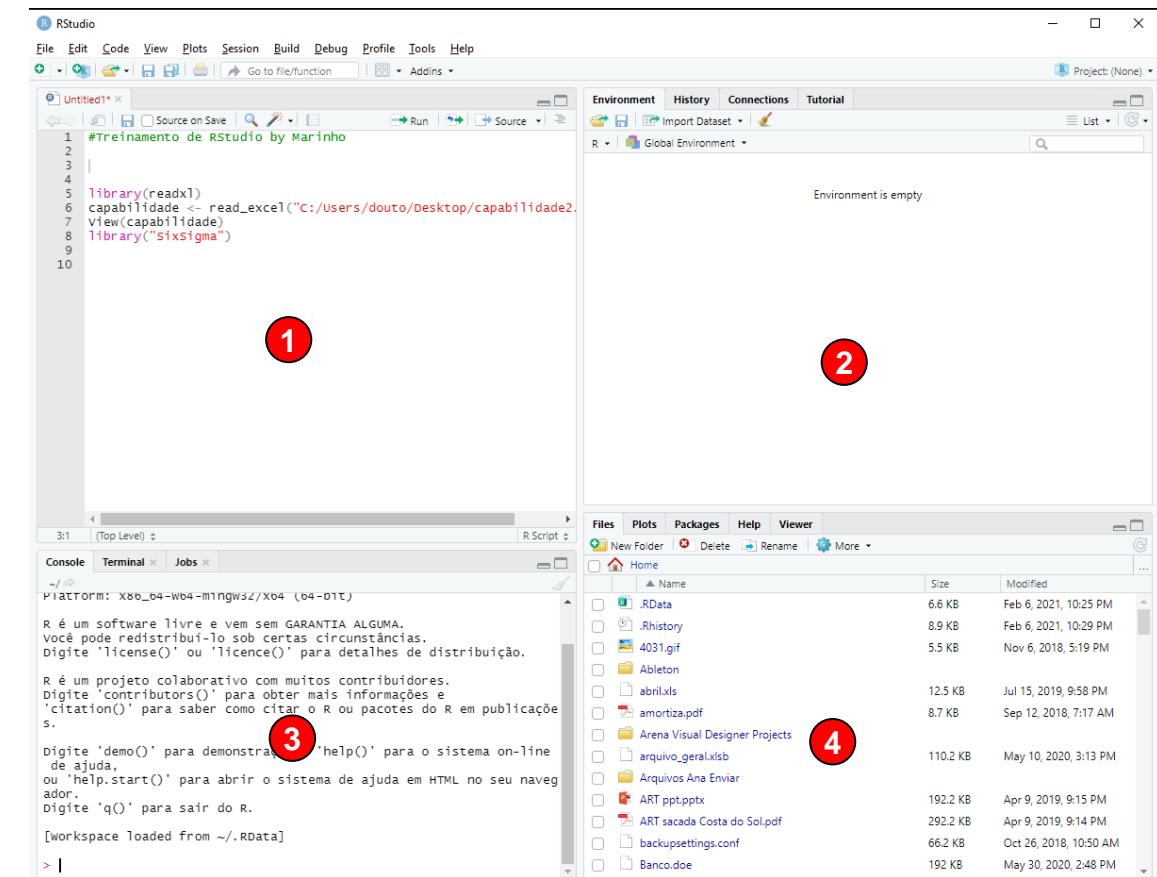
Resumidamente, os principais componentes do software estão apontados ao lado :

1 – Script: Local onde são digitados os comandos necessário para realizar as análises. Ele é chamado de script pois, é uma sequência de comandos que, aplicados a diferentes dados, realizarão as mesmas etapas de análises.

2 – Ambiente: Local onde são apresentadas as variáveis de ambiente, bem como históricos e conexões a bancos de dados

3 – Console e Terminal: Todos os comandos realizados no Script serão mostrados no console, que é a saída ou resposta para aquele comando realizado.

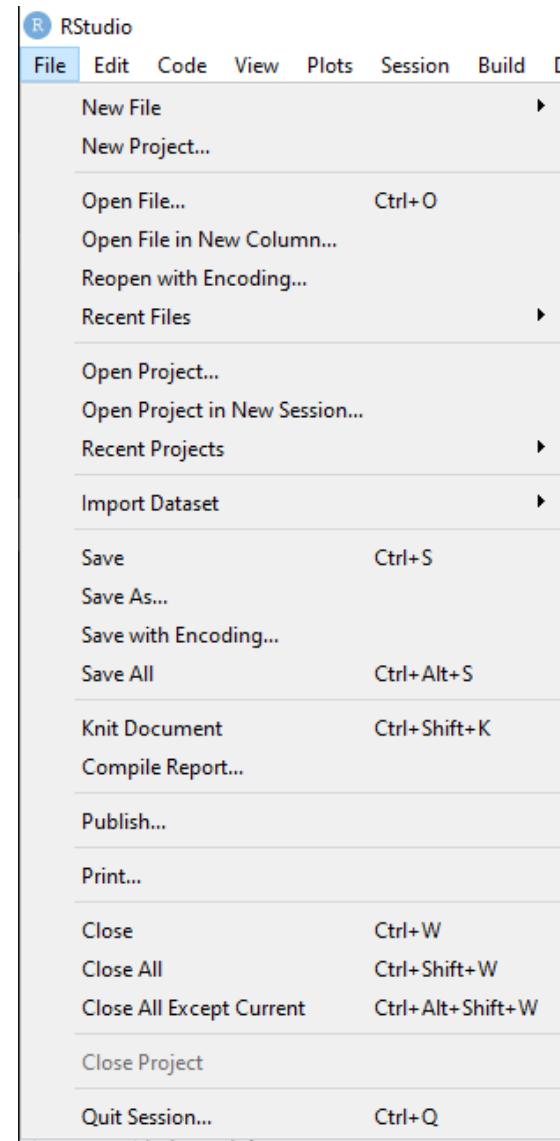
4 – Navegador: Local onde pode-se navegar pelos arquivos do projeto, pelos gráficos que foram gerados, pelos pacotes e bibliotecas que foram carregados, bem como instalar novos pacotes, pela ajuda e pelo visualizador.



File

O ícone “File” deve ser utilizado quando se deseja :

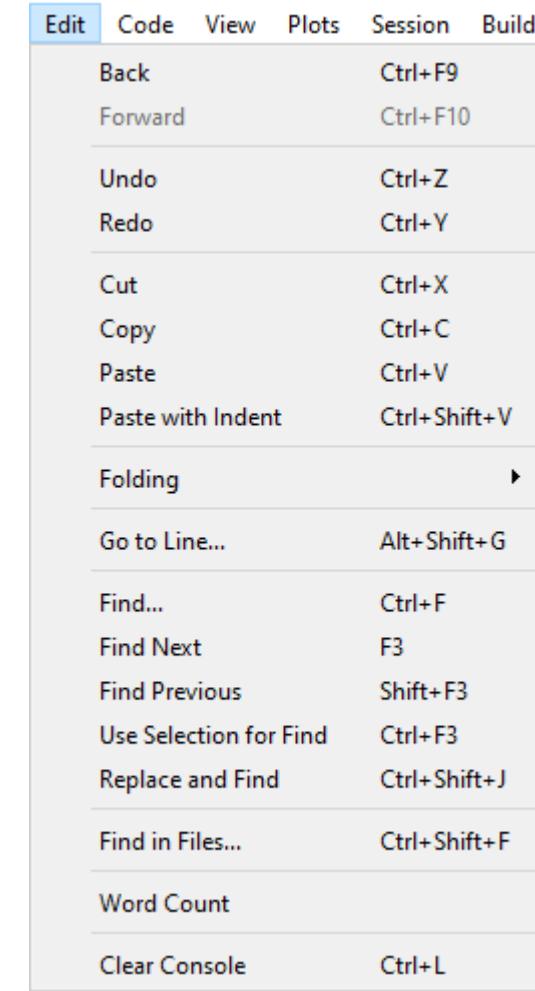
- Criar um novo arquivo ou projeto
- Abrir um arquivo ou projeto
- Importar dados de outras fontes
- Salvar seu projeto
- Publicar seu projeto
- Fechar seu projeto
- Sair do programa



Edit

O ícone “Edit” deve ser utilizado quando se deseja :

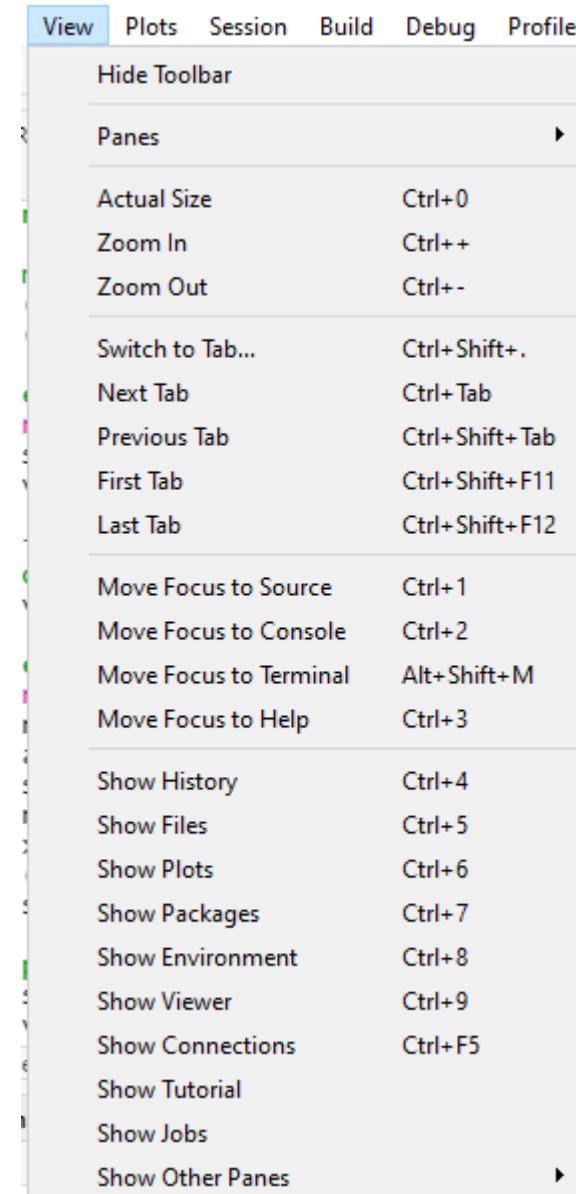
- Desfazer o último comando executado
- Cancelar o comando “Desfazer”
- Procurar um texto no script
- Ir para determinada linha de programação
- Substituir texto no script
- Encontrar nos arquivos
- Contar palavras
- Limpar o console



View

O ícone “View” deve ser utilizado quando se deseja :

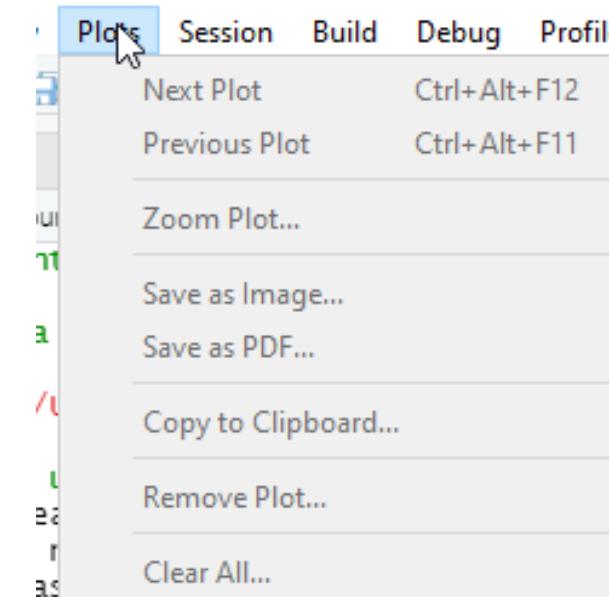
- Habilitar ou desabilitar painéis
- Dar o tirar zoom das telas
- Mudar de abas nos painéis
- Mover o foco do cursor
- Mostrar ou ocultar painéis como:
 - Histórico
 - Arquivos
 - Pacotes
 - Ambiente
 - Conexões, entre outros



Plots

O ícone “Plots” deve ser utilizado quando se deseja :

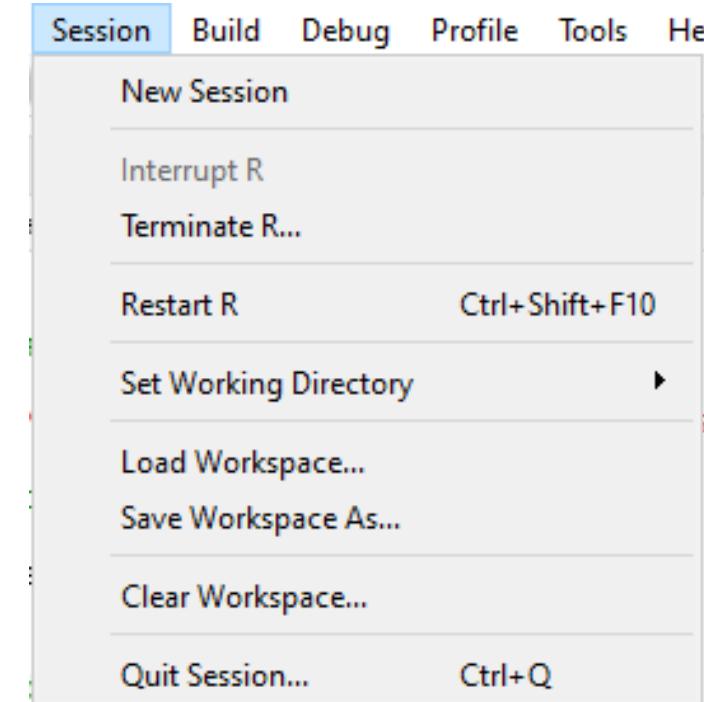
- Visualizar os gráficos plotados durante o projeto, avançando ou retornando
- Dar zoom no gráfico
- Salvar em imagem ou pdf
- Copiar para a área de transferência
- Remover o gráfico
- Apagar todos os gráficos



Session

O ícone “Session” deve ser utilizado quando se deseja realizar :

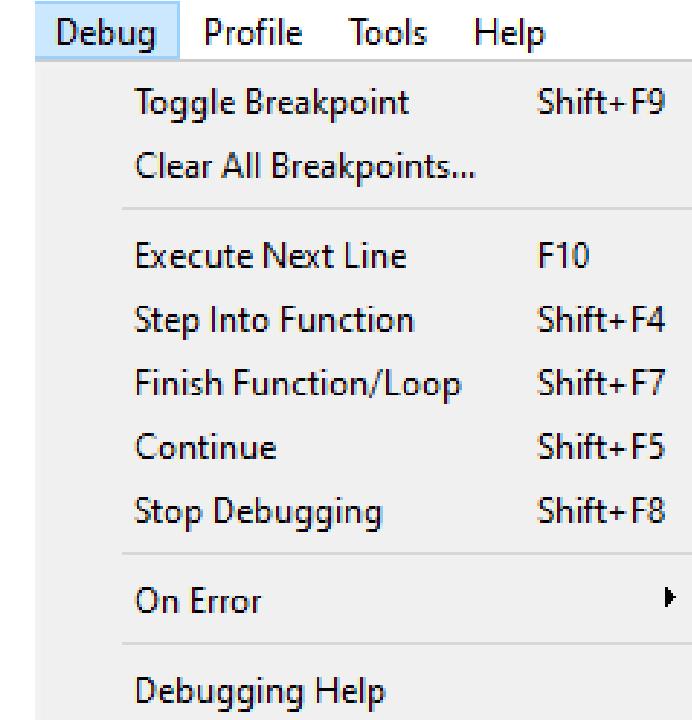
- A abertura de uma nova sessão
- Interromper um comando que está rodando
- Finalizar um comando que está rodando
- Reiniciar o interpretador R
- Setar o diretório de trabalho
- Carregar ou Salvar uma área de trabalho
- Limpar a área de trabalho
- Finalizar a sessão.



Debug

O ícone “Debug” deve ser utilizado quando se deseja realizar :

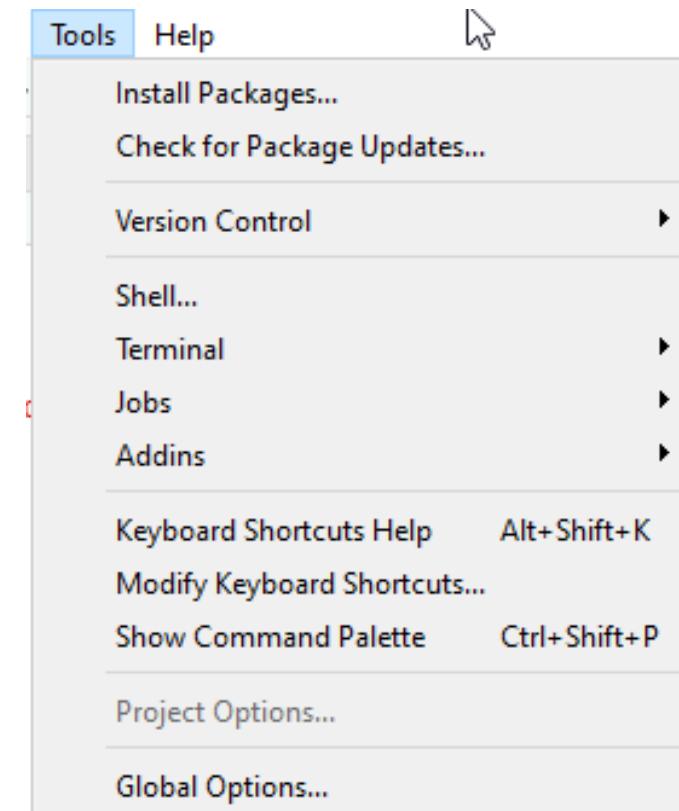
- Análises (depuração) do Código com
 - Criar ou limpar pontos de parada
 - Executar próxima linha
 - Entrar em uma função
 - Finalizar um loop de execução
 - Continuar uma interrupção
 - Parar a depuração
 - Definir que ação tomar em caso de erros
 - Entrar no help do depurador.



Tools

O ícone “Tools” deve ser utilizado quando se deseja :

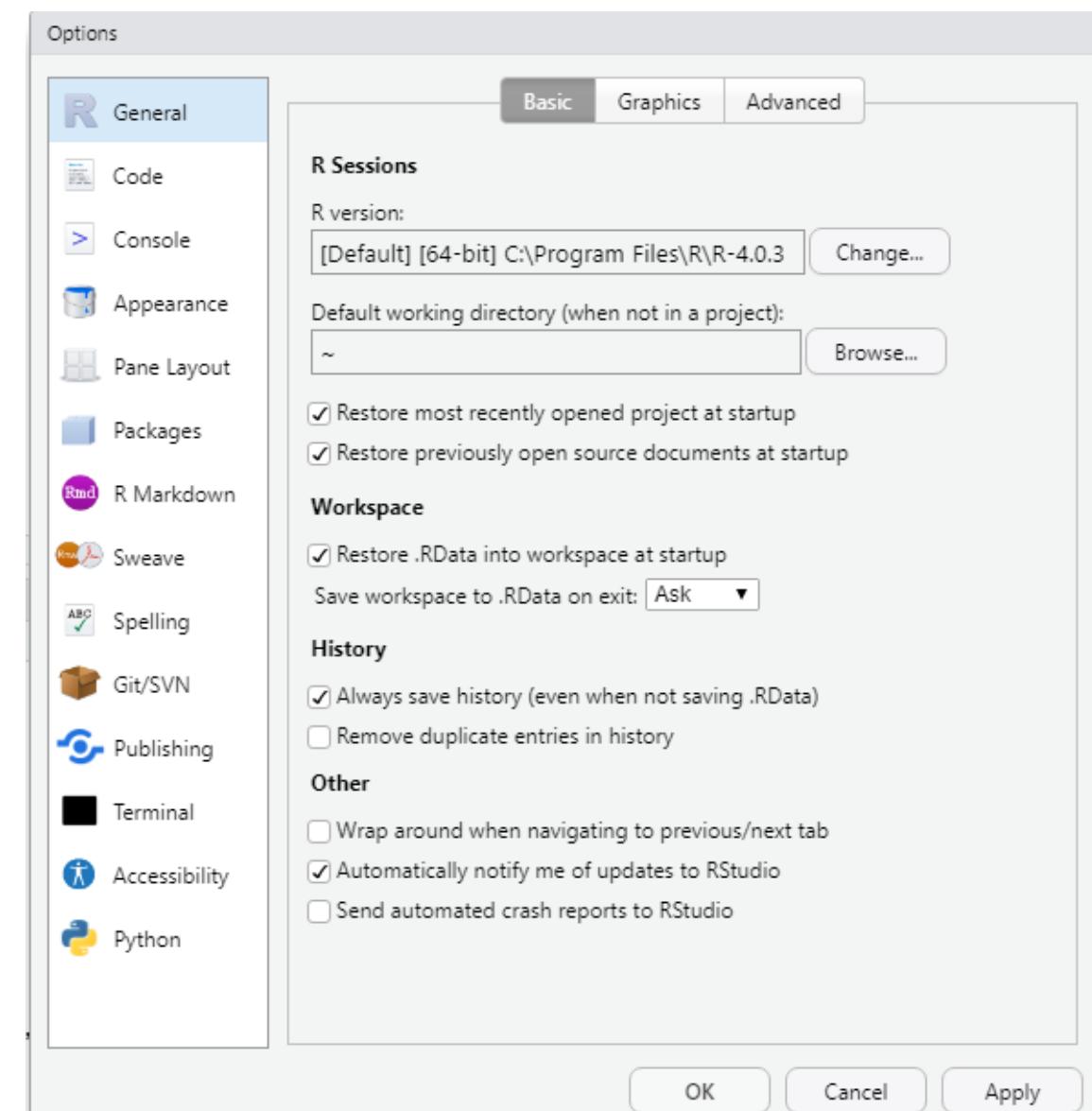
- Instalar um novo pacote
- Verificar atualizações de pacotes
- Criar um controle de versão do seu script
- Abrir um shell ou terminal
- Habilitar plugins
- Visualizar atalhos de teclado
- Modificar os atalhos
- Mostrar a paleta de comandos
- Ver opções do projeto
- Abrir as configurações do RStudio



Options

As configurações do Rstudio são diversas. Não entraremos em detalhes pois é muita coisa, as aqui podemos:

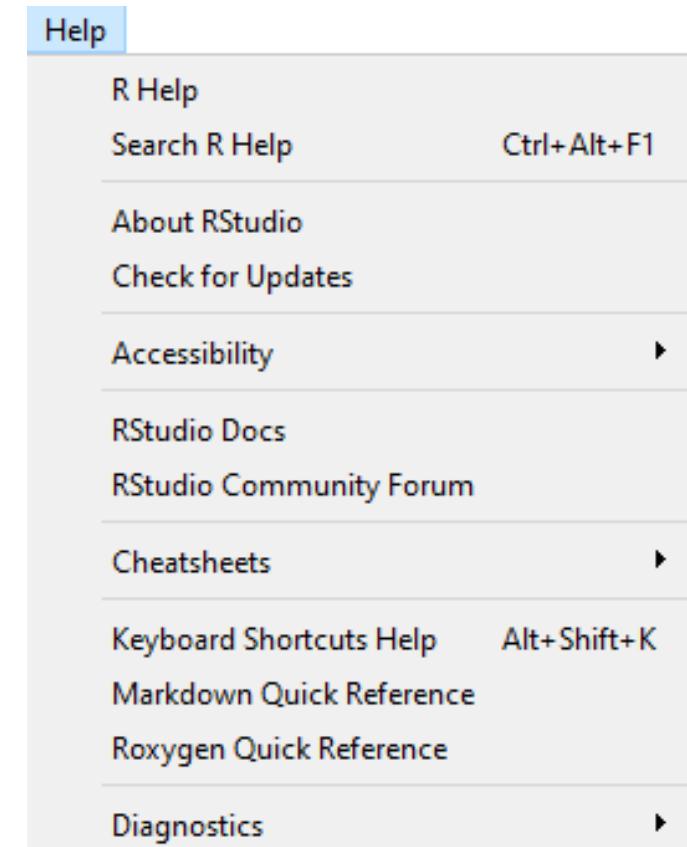
- Escolher a versão do R a ser usada
- Configurar as telas de código e console
- Mudar a aparência da tela, inclusive deixar em modo noturno
- Visualizar os pacotes
- Configurar correção ortográfica
- Configurar outras opções de ambiente
- Configurar interação com Python.



Help

O ícone “Help” deve ser utilizado quando se deseja:

- Abrir o navegador na página de help do R
- Ver a versão do seu Rstudio
- Checar por updates do Rstudio
- Abrir o navegador na página de help do Rstudio
- Abrir os Cheatsheets (dicas para ficar mais ágil no uso do programa)
- Ver ajuda dos atalhos de teclado
- Realizar um diagnóstico, caso seu Rstudio esteja se comportando de maneira anômala.

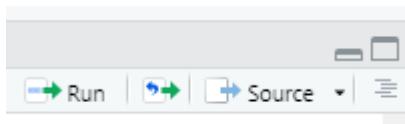
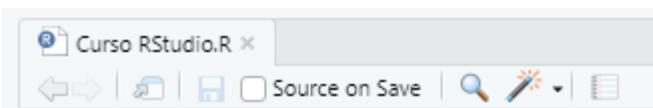


Abas dos painéis

Quase tudo que existe nos menus, vistos anteriormente, pode ser feito diretamente na tela do Rstudio.

Na tela 1: temos o nome do script e botões para salvar ou fechar. Mais à direta, botão para rodar o programa ou rodar linha a linha. Há também botões para maximizar ou minimizar esse painel.

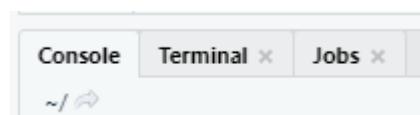
Caso você use o comando `View()`, a saída dele também será mostrado aqui no formato de uma aba



No segundo painel, abaixo, temos 3 abas: a primeira dá a saída de todos os comandos realizados acima. É a resposta do seu pedido.

No terminal pode-se abrir um bash e dar um comando para o sistema.

E por final a aba Jobs irá mostrar os trabalhos que estão rodando em background (plano de fundo).

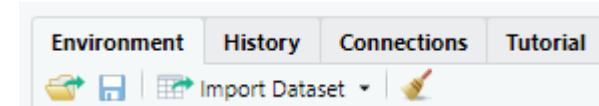


Abas dos painéis

No terceiro painel temos 4 abas. Environment mostra as variáveis de processo, todos os objetos do projeto são mostrados aqui. Pode-se também abrir uma planilha.

Na aba history fica registrado todos os comando já executados, um histórico. Já em conexões, caso você esteja se comunicando com algum banco de dados

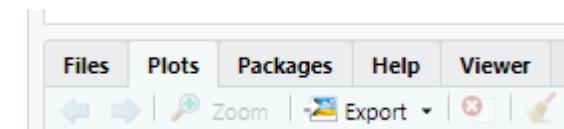
Na aba tutorial, pode ser aberto um tutorial do RStudio



No ultimo painel temos 5 abas. A primeira é um visualizador da árvore de arquivos do projeto. Muito útil para definir o diretório de trabalho

Temos a aba Plots onde aparecerão os gráficos que fizermos. Podemos também salvar ou gerar PDF. No Packages, podemos ver os pacotes instalados, carregar pacotes e instalar pacotes.

Na aba help, todo help pedido no script sairá ali. E por fim o Viewer é um visualizador que nunca usei para nada.



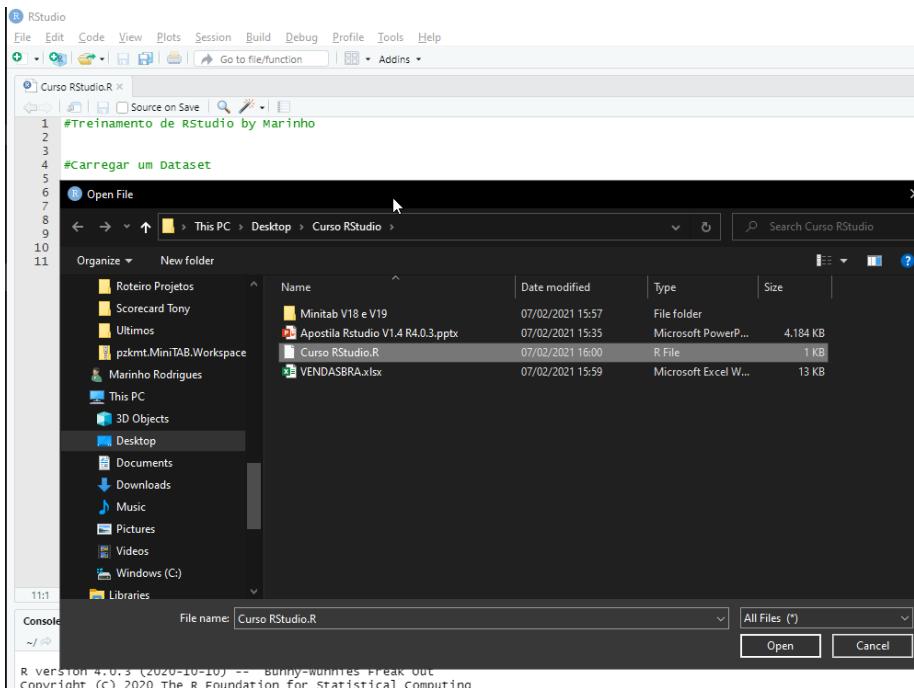


Entrada e Manipulação de Dados

Nível: Básico

Exercício 1 - Abrindo um Script Salvo

- 1 Abra o RStudio
- 2 Escolha **File ▶ Open File**
- 3 Selecione **Curso RStudio.R**



- 4 Clique em **Open**

Carregando um Dataset

Use Ctrl+Enter em cada linha para rodar o script
Linhas que começam com # são meros comentários.
Nota: Altere o script para o endereço em seu computador

```
#Treinamento de RStudio by Marinho
#Carregar um Dataset
getwd()
setwd("C:/Users/douto/Desktop/Curso RStudio/Datasets")

library(readxl)
vendas <- read_excel("VENDASBRA.xlsx")
View(vendas)
```

A saída no console será algo do tipo

```
> getwd()
[1] "C:/Users/douto/Documents"
> setwd("C:/Users/douto/Desktop/Curso RStudio/Datasets")
> library(readxl)
> vendas <- read_excel("VENDASBRA.xlsx")
> View(vendas)
>
```

E ao lado do Script, será apresentada a tabela com os dados.

	Date	Vendas Caçapava	Vendas Camaçari	Vendas Betim	Vendas Gravataí	Month	Quarter
1	2005-01-01	1436	1524	1640	1453	Jan2005	Q12005
2	2005-01-02	1467	1442	1599	1451	Jan2005	Q12005
3	2005-01-03	1521	1539	1797	1449	Jan2005	Q12005
4	2005-01-04	1561	1483	1687	1454	Jan2005	Q12005
5	2005-01-05	1549	1596	1744	1456	Jan2005	Q12005
6	2005-01-06	1489	1477	1618	1451	Jan2005	Q12005
7	2005-01-07	1489	1499	1630	1451	Jan2005	Q12005
8	2005-01-08	1536	1455	1737	1446	Jan2005	Q12005

Explorando os Dados

Decidiu-se analisar os dados de Betim gerando estatísticas descritivas e sumários gráficos.

Comece a rodar o Script, usando Ctrl+Enter em cada linha:

```
#Treinamento de RStudio by Marinho

#Carregar um Dataset

getwd()
setwd("C:/users/douto/Desktop/Curso RStudio/Datasets")

library(readxl)
vendas <- read_excel("VENDASBRA.xlsx")
view(vendas)
```

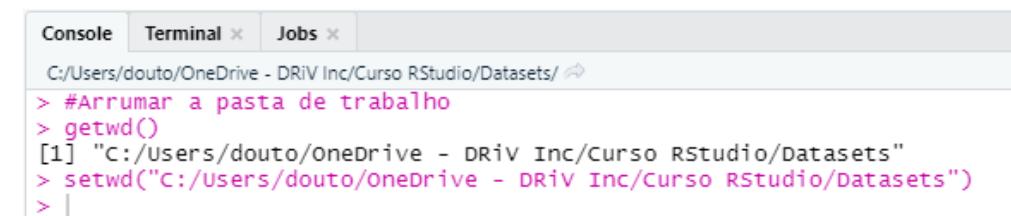
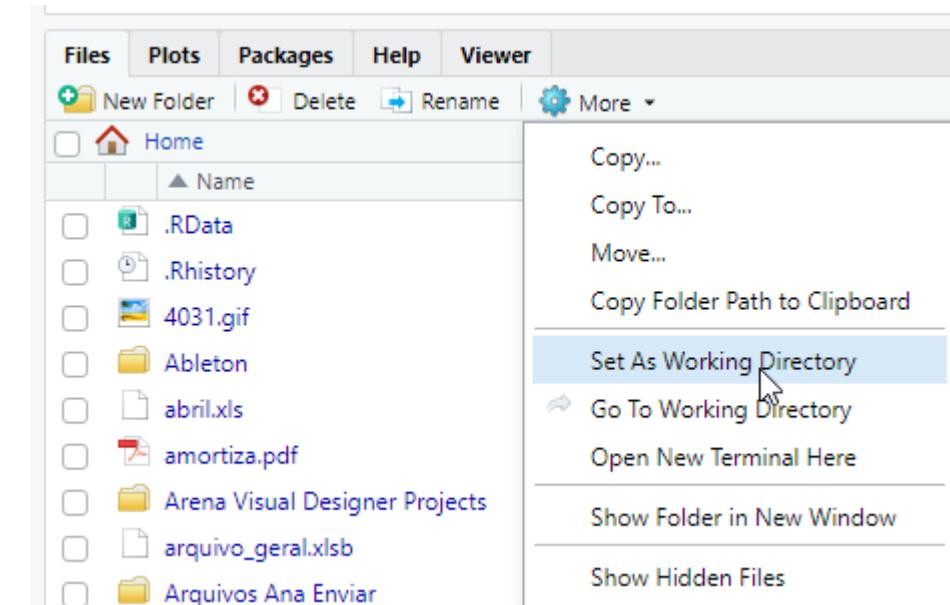
As linhas com # no começo são comentários, elas não rodam nada.

getwd() significa: get the working directory e mostra a pasta atual de trabalho. Normalmente seus dados não estão nela, assim digite:

setwd("c:/Users/<seu usuário>/<sua pasta...>") para definir onde estão seus dados.

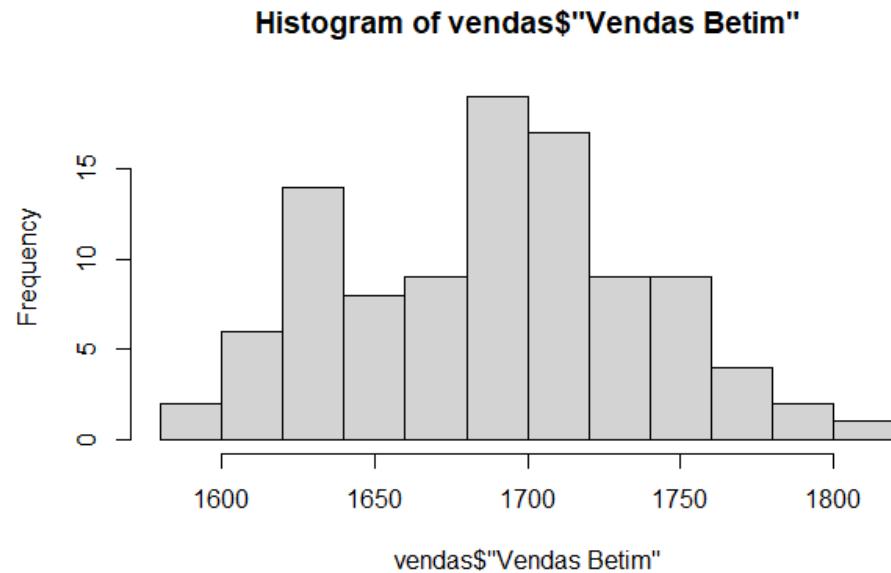
Ou vá na aba "files" do quarto painel, navegue até a pasta necessária e defina como pasta de trabalho

Interpretando os Resultados



Analisando os dados Importados

Uma boa maneira de se avaliar o centro, a dispersão e a forma geral de uma distribuição, é elaborar um histograma, como o disposto abaixo :



O histograma de Vendas Betim, mostra que a distribuição é relativamente simétrica (provavelmente normal), com a maioria das observações no centro e poucas observações nas extremidades.

Procedimento:

Para solicitar um histograma com a curva normal, devemos seguir os procedimentos abaixo.

```
10 library(readxl)
11 vendas <- read_excel("VENDASBRA.xlsx")
12 view(vendas)
13
14 #Analisando os dados
15 hist(vendas$'Vendas Betim', nclass = 15)|
```

Note que o Dataset está armazenado na variável <vendas>

O comando para gerar o histograma é o hist()

Para chamar a coluna “Vendas Betim” use o operador \$ e como há um espaço entre Vendas e Betim, é necessário colocar as aspas simples. Caso não houvesse espaço isso não seria necessário.

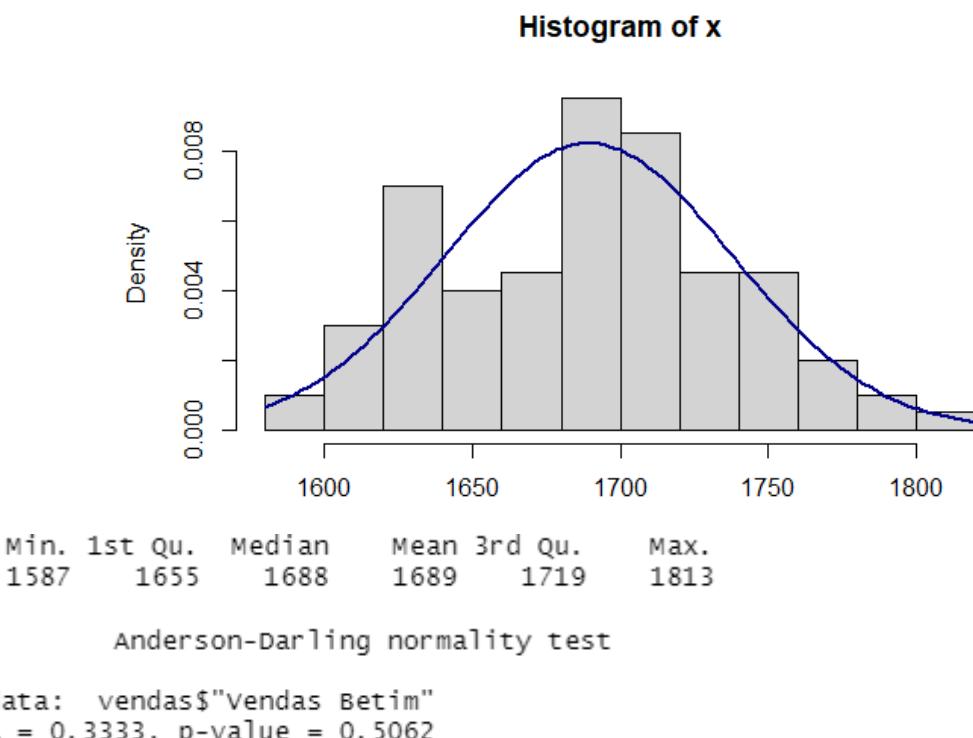
Logo: vendas\$'Vendas Betim' seleciona a coluna do dataset.

Por fim, nclass = 15 pede para que o gráfico tenha 15 barras

Testando a Normalidade

O sumário gráfico de **Display Descriptive Statistics** inclui:

- um histograma com curva normal
- intervalos de confiança
- testes de normalidade
- estatísticas descritivas dos dados



Exibir Estatísticas Descritivas

```
18 #Teste de normalidade Anderson-Darling
19 library("nortest")
20 summary(vendas$"Vendas Betim")
21 m<-mean(vendas$"Vendas Betim")
22 std<-sqrt(var(vendas$"Vendas Betim"))
23 x<-vendas$"Vendas Betim"
24 hist(x, probability = T, nclass = 15)
25 curve(dnorm(x, mean=m, sd=std), col="darkblue", lwd=2, add=TRUE)
26 ad.test(vendas$"Vendas Betim")
27 |
```

Carregar a biblioteca nortest

Com o comando Summary Podemos ter as estatísticas descritivas de Vendas Betim

Depois, para facilitar os próximos comandos, vamos armazenar os dados em variáveis:

m = será a média

std = o desvio padrão

Em x, colocaremos os valores da coluna betim

Rodamos novamente o histograma, depois a PDF (Probability Distribution Function) que é a curva azul.

Por fim, pedimos o teste de normalidade que resulta em nos dizer que essa é com certeza uma curva normal

Exercício 2 – Manipulando Dados

Problema

Os dados de vendas importados para o RStudio foram digitados em linhas e precisam estar no formato de coluna.

Coleta de Dados

Foram coletados dados mensais das receitas de vendas de janeiro a maio.

```
28 #Manipulando dados
29 vendas2<- read_excel("VENDAS2.xlsx")
30 View(vendas2)
31
```

	Região	Janeiro	Fevereiro	Março	Abril	Maio
1	Escobar	387980	451700	456789	435678	349050
2	Gravataí	578990	600987	456789	781656	456798
3	Caçapava	435800	542700	345988	345678	564050
4	Camaçari	497050	827900	456789	897658	687050
5	Betim	613242	611689	456789	560988	434567

Organizar dados

```
28 #Manipulando dados
29 vendas2<- read_excel("VENDAS2.xlsx")
30 view(vendas2)
31 library("reshape2")
32 vendas2a<- melt(vendas2, id.vars = 1)
33 view(vendas2a)
34 colnames(vendas2a)<- c("Região", "Mês", "Vendas")
35
```

Seguindo na linha 31, carregamos uma biblioteca chamada reshape2

Usamos a função melt para reagrupar os dados e armazenamos na variável vendas2a

Por fim, para mudar os nomes das colunas ao lado usamos o comando da linha 34.

Assim deixamos os dados de forma estruturada para futuras análises.

	Região	variable	value
1	Escobar	Janeiro	387980
2	Gravataí	Janeiro	578990
3	Caçapava	Janeiro	435800
4	Camaçari	Janeiro	497050
5	Betim	Janeiro	613242
6	Escobar	Fevereiro	451700
7	Gravataí	Fevereiro	600987
8	Caçapava	Fevereiro	542700
9	Camaçari	Fevereiro	827900
10	Betim	Fevereiro	611689
11	Escobar	Março	456789
12	Gravataí	Março	456789
13	Caçapava	Março	345988
14	Camaçari	Março	456789
15	Betim	Março	456789
16	Escobar	Abril	435678
17	Gravataí	Abril	781656
18	Caçapava	Abril	345678
19	Camaçari	Abril	897658
20	Betim	Abril	560988
21	Escobar	Maior	340000



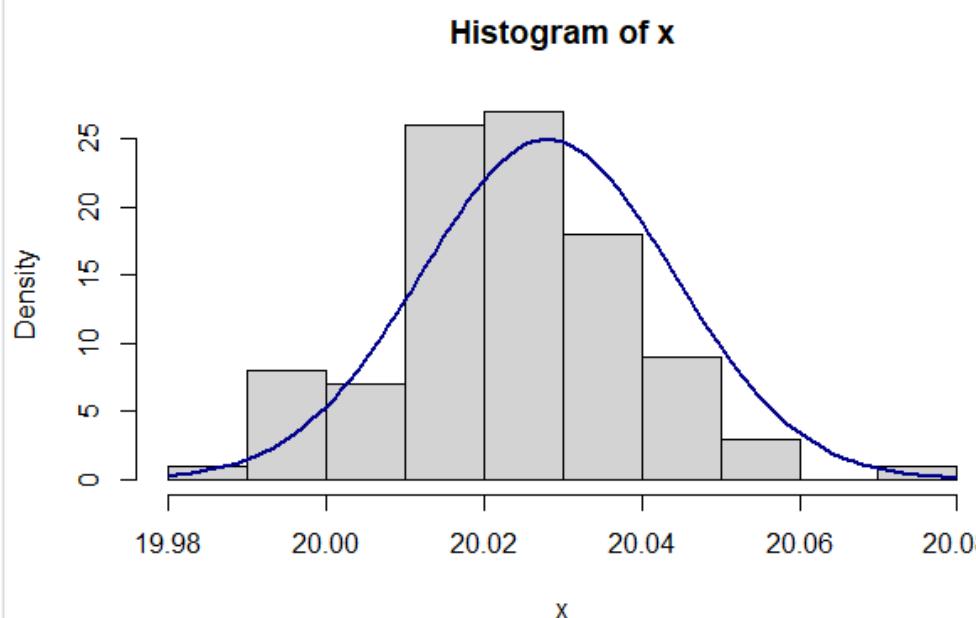
Análise Gráfica

Nível: Básico

A) Sumário

Sempre que criarmos um dataset, temos que entender os números, para isso, devemos fazer o seguinte ritual:

Arquivo :
Grampo C.xlsx



Exibir Estatísticas Descritivas

```

36 #Analisando um dataset
37 grampo<- read_excel("GRAMPO C.xlsx")
38 show(grampo)
39 x<- grampo$`Dim Interna`
40 m<- mean(grampo$`Dim Interna`)
41 std<- sqrt(var(grampo$`Dim Interna`))
42 summary(x)
43 hist(x, probability = T, nclass = 10)
44 curve(dnorm(x, mean=m, sd=std), col="darkblue", lwd=2, add=TRUE)
45 ad.test(x)
46 boxplot(x, horizontal = TRUE)

```

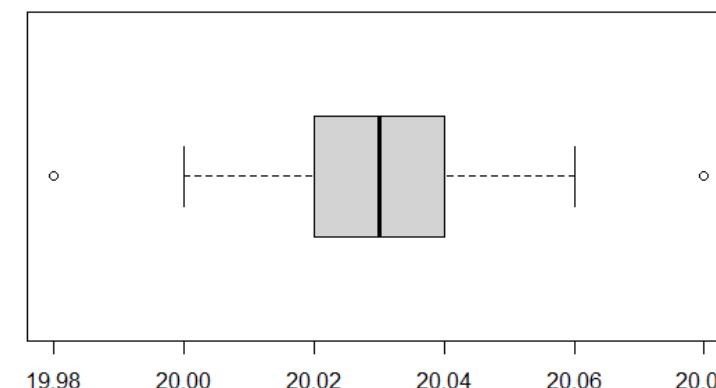
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	19.98	20.02	20.03	20.03	20.04	20.08

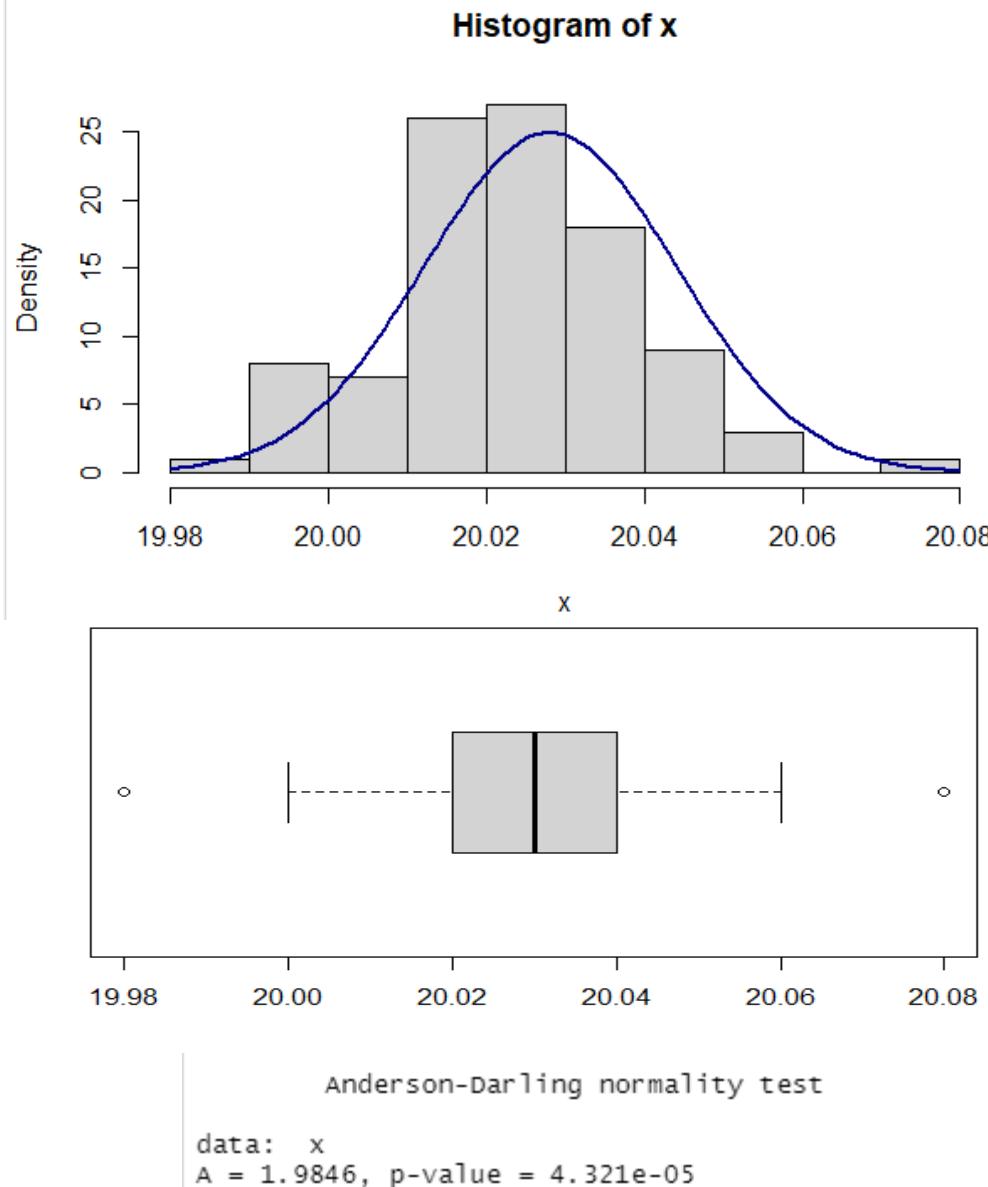
Anderson-Darling normality test

```

data: x
A = 1.9846, p-value = 4.321e-05

```





Teste de normalidade Anderson-Darling

Analizando a curva do exemplo acima, temos uma idéia aproximada que a distribuição é normal.

No sumário gráfico, é possível avaliar se a distribuição é normal ou não, através do teste de normalidade Anderson-Darling, o qual ajuda a avaliar 2 hipóteses opostas :

- A *hipótese nula* (H_0) : a população é normal
- A *hipótese alternativa* (H_a) : a população não é normal.

O valor-p (p-value) é a probabilidade de se obter uma amostra não normal. Em Six Sigma, tipicamente utilizamos um α (alfa, valor-p) igual a 0,05.

- Se o **p-value** for $<$ ou $=$ a α , rejeitamos H_0
- Se o **p-value** for $>$ que α , não rejeitamos H_0 (nunca aceitamos H_0 , simplesmente não rejeitamos).

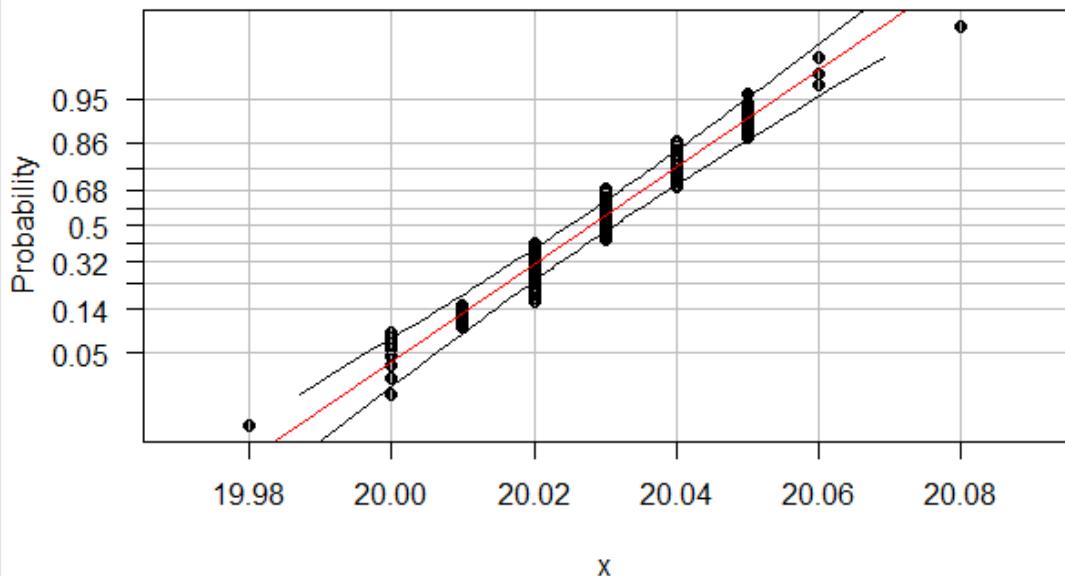
No exemplo acima, o p-value é $<0,005$. Com isto, rejeitamos a hipótese nula, ou seja, podemos afirmar que a população não é normal (ao contrário do que pensávamos inicialmente).

B) Gráfico de Probabilidade Normal

Outra forma de analisar a normalidade de um Dataset é pelo gráfico da probabilidade normal

Nele, todos os pontos devem estar dentro de uma determinada região, formada por uma linha, que é a própria distribuição normal.

Probability Plot for "normal" distribution



Para isso, basta adicionar uma nova biblioteca e chamar a função ppPlot.

Note que a cada linha de comando aparecem coisas novas ou novas formas de se fazer a mesma coisa. Isso é proposital para que você aprenda a usar o programa.

Antes de carregar a biblioteca, vamos instalá-la. E então testarmos a normal.

```
48 install.packages("qualityTools")
49 library("qualityTools")
50 ppPlot(x, "normal")
51 help(ppPlot)
```

Note que na linha 55 testamos se os dados são normais, mas podemos fazer outros testes como log-normal, exponencial, Weibull, logística entre outros, para isso basta buscar o help.



Probability Plots for various distributions

- "beta"
- "cauchy"
- "chi-squared"
- "exponential"
- "f"
- "gamma"
- "geometric"
- "log-normal"
- "lognormal"
- "logistic"
- "negative binomial"
- "normal"
- "Poisson"
- "t"
- "weibull"

C) Gráfico de Séries Temporais (TS Plot)

Para criar um gráfico que mostra alguma informação ao longo do tempo, podemos usar a seguinte função:

```
53 #Séries Temporais
54 inspec<- read_excel("INSPEC.xlsx")
55 view(inspec)
56 inspec.timeseries<- ts(inspec$`Tempo Inspeção`)
57 plot(inspec.timeseries)
```

Carregamos o arquivo INSPEC.xlsx na variável inspec

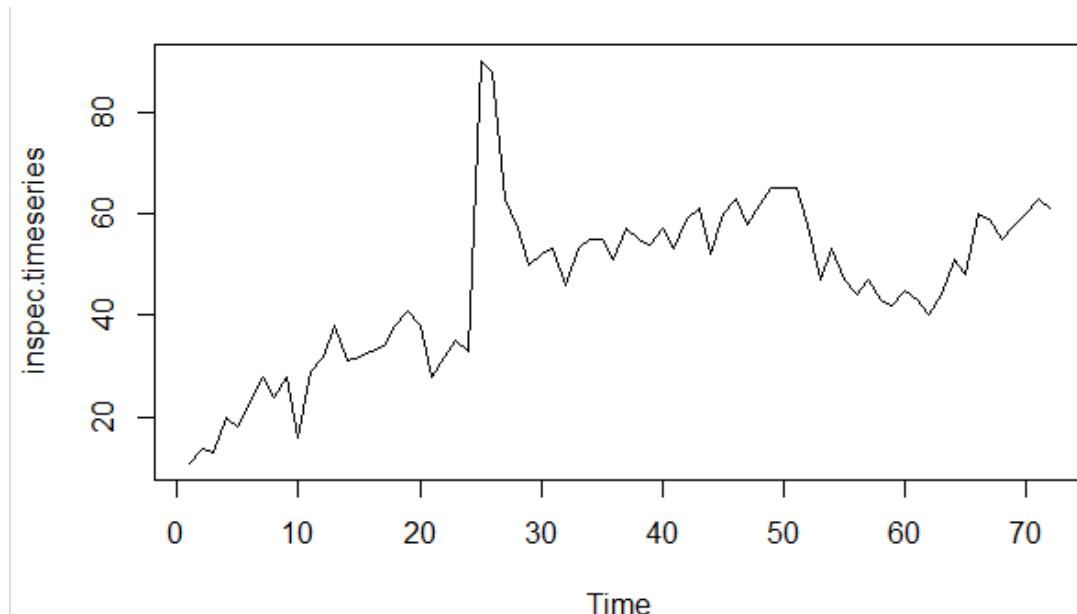
Depois precisamos colocar os dados da forma adequada para fazer o gráfico, com a função TS e gravamos na variável inspec.timeseries

Depois plotamos o gráfico

Interpretação dos Resultados

O gráfico mostra que os tempos de inspeção do grampo C aumentam ao longo do dia e também ao longo de vários dias de trabalho, e apresentam uma certa estabilidade do meio do gráfico para a frente.

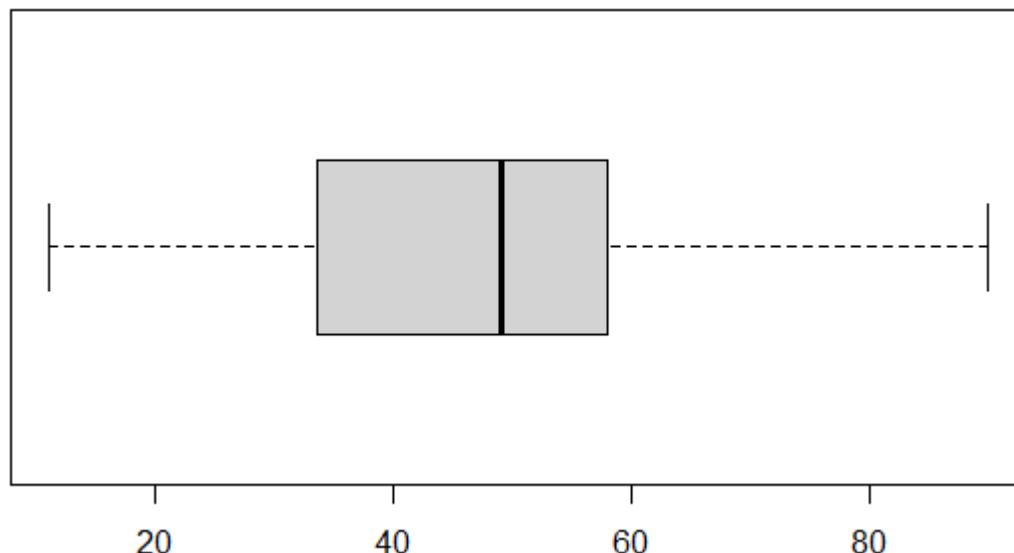
Podemos perceber ainda, que existe um grande aumento, porém de curta duração (no meio do gráfico).



D) Gráfico de Caixa (Box Plot)

Já fizemos um boxplot antes, mas vamos recapitular

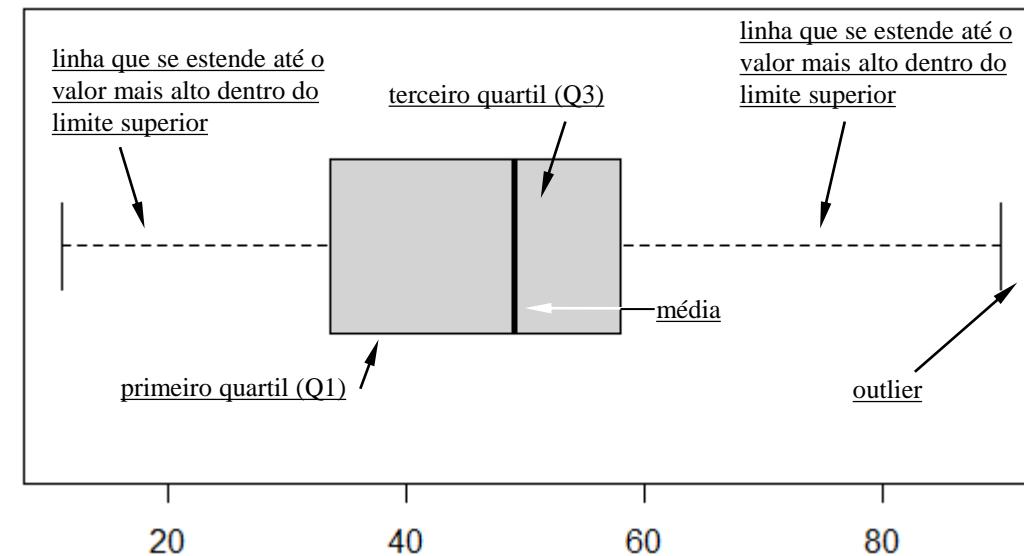
```
59 boxplot(inspec$`Tempo Inspeção`, horizontal = TRUE)
60 summary(inspec$`Tempo Inspeção`)
```



	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	11.00	33.75	49.00	46.33	58.00	90.00

Interpretação dos Resultados

No caso abaixo as linhas acima e abaixo da caixa possuem o mesmo tamanho. Quando uma linha é mais longa que a outra , temos a indicação de que os dados não são simétricos.



Boxplots são particularmente úteis para comparar várias amostras. Pode-se exibir várias caixas lado a lado e comparar suas distribuições com respeito à simetria, variabilidade e tendência central.



Estudos R&R de Variáveis

Nível: Intermediário

Análise do sistema de medição (variável)

O que é análise do sistema de medição ?

Ao medir a saída de um processo, depara-se com duas fontes de variação:

- Variação do processo produtivo
- Variação do sistema de medição

Caso a variação do sistema de medição seja grande, em comparação com a variação do processo produtivo, a medição tomada será de pouco valor.

Quando utilizar análise dos sistemas de medição ?

O sistema de medição deve discriminar de forma adequada as diferentes peças, de modo a determinar o controle e a capacidade do processo. Utilize a análise dos sistemas de medição antes de analisar o controle e a capacidade do processo para confirmar se o sistema de medição está executando as medidas de forma consistente e acurada e se ele é capaz de distinguir as peças umas das outras de forma adequada.

Por que usar análise dos sistemas de medição ?

Utilize a análise dos sistemas de medição quando desejar obter respostas para os seguintes tipos de questões :

- O sistema de medição é capaz de distinguir as peças uma das outras de forma adequada ?
- O sistema de medição é estável ao longo do tempo ?
- O sistema de medição é acurado para todas as peças envolvidas no processo ?

Por exemplo :

- O viscosímetro é capaz de distinguir adequadamente a diferença de viscosidade de várias amostras de tinta ?
- A balança precisa ser recalibrada periodicamente para pesar embalagens de batata chips de forma acurada ?
- O termômetro mede acuradamente a temperatura de todas as graduações de calor utilizadas no processo ?

Qualidade das medidas

A qualidade das medidas podem ser classificadas em duas categorias : acurácia e precisão.

- A acurácia descreve a diferença entre a medição e o valor real da peça;
- A precisão descreve a variação que se percebe ao medir a mesma peça repetidamente com o mesmo instrumento de medição.

É possível haver a ocorrência de um desses problemas dentro de um sistema de medição ou de ambos. Por exemplo, pode-se ter um instrumento que meça peças com precisão (pouca variação nas medições), mas não com acurácia. Pode-se também ter um instrumento que é acurado (a média das medições é muito próxima ao valor alvo), mas não é preciso. Ou seja, há grande variação entre as medições. Também pode-se ter um instrumento que não é acurado ou preciso.



Acurácia

A acurácia de um sistema de medição é geralmente dividida em três componentes :

- linearidade – é uma medida de como o tamanho da peça afeta o vício do sistema de medição. É a diferença nos valores do vício observados ao longo da *faixa natural* do processo;
- vício – é a diferença entre a média dos valores medidos com o instrumento a ser avaliado e um *valor de referência*;
- estabilidade – é uma medida da qualidade do desempenho do sistema de medição ao longo do tempo. É a variação total obtida com um instrumento em particular, com a mesma peça, ao se medir uma única característica *ao longo do tempo*.

Precisão

A precisão, ou variação da medição, pode ser dividida em dois

componentes :

- repetibilidade – é a variação observada quando o mesmo operador mede a mesma peça repetidas vezes com o mesmo instrumento de medição.
- reprodutibilidade – é a variação observada quando diferentes operadores medem a mesma peça usando o mesmo instrumento de medição.

Determinação do sistema de medição

Desejamos determinar :

- A capacidade do sistema de medição de distinguir as peças umas das outras;
- Se os operadores são capazes de usar o sistema de forma consistente.

Para responder a essas questões utilize a opção Gage R&R Study (Crossed).

Tolerância

Os limites de especificação do torque do parafuso de fixação do suporte do cinto de segurança são 50 ± 5 Nm. Em outras palavras, o torque pode variar até 5 Nm em qualquer direção. A tolerância é a diferença entre os limites superior e inferior de especificação; aqui, $55 - 45 = 10$ Nm.

Ao digitar o valor para uma tolerância, pode-se determinar qual proporção da faixa de tolerância será usada pela variação do sistema de medição.

Estudo de R&R (cruzado)

Siga o script e note a ordem do ensaio:

```
62 #Gage R&R
63 #nota: o arquivo tem que ser feito assim:
64 #operador | Peça | Repetição
65 # 1 1 1
66 # 2 1 1
67 # 1 2 1
68 # 2 2 1
69 #...
70 # 1 1 2
71 # 2 1 2
72 # 1 2 2
73 # 2 2 2
74 ...
75 torque<- read_excel("TORQUE.xlsx")
76 gdo = gageRRDesign(2, 10, 2, method = "crossed", randomize = FALSE)
77 response(gdo) = torque$Torque
78 gdo = gageRR(gdo, tolerance = 10)
79 summary(gdo)
80 plot(gdo)
```

A	B	C	D	E	F	G	H	I	J
1 Operador	Banco	Repet	Torque						
2 1	1	1	40,5						
3 2	1	1	40,2						
Sort									
6	Add Level	Delete Level	Copy Level	▲	▼	Options...	?	X	My data has headers
7	Column	Sort On	Order						
8 Sort by	Repet	Cell Values	Smallest to Largest						
9 Then by	Banco	Cell Values	Smallest to Largest						
10 Then by	Operador	Cell Values	Smallest to Largest						
11									
12									

Interpretação dos resultados

Tabelas da análise de variância

O RStudio utiliza análise de variância (ANOVA) para calcular os componentes da variação, que são usados em seguida para estimar a variação percentual causada pelo sistema de medição. A variação percentual aparece na tabela Gage R&R.

A tabela ANOVA para 2 fatores inclui termos para peças (**Banco**), operador (**Operador**) e interação operador e peça (**Operador*Banco**).

Se o valor p da interação entre os 2 fatores (Operador e peça) for $\geq 0,25$, isso significa que existe interação entre os operadores.

Aqui, o valor p para **Operador*Banco** é de 0,109. Portanto, o RStudio não gerou uma segunda tabela ANOVA sem o termo da interação (neste exemplo, a interação entre operador e banco é importante para o estudo do sistema de medição).

Operators:	2	Parts:	10			
Measurements:	2	Total:	40			
<hr/>						
Anova Table - crossed Design						
Df	Sum Sq	Mean Sq	F value	Pr(>F)		
Operator	1	0.00	0.000	0.006 0.938		
Part	9	65.07	7.230	177.429 <2e-16 ***		
Operator:Part	9	0.70	0.078	1.915 0.109		
Residuals	20	0.81	0.041			
<hr/>						
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						
<hr/>						
Gage R&R						
	VarComp	VarCompContrib	stdev	studyVar	studyVarContrib	P/T Ratio
totalRR	0.0594	0.0321	0.244	1.462	0.179	0.1462
repeatability	0.0407	0.0221	0.202	1.211	0.149	0.1211
reproducibility	0.0186	0.0101	0.137	0.819	0.100	0.0819
Operator	0.0000	0.0000	0.000	0.000	0.000	0.0000
Operator:Part	0.0186	0.0101	0.137	0.819	0.100	0.0819
Part to Part	1.7881	0.9679	1.337	8.023	0.984	0.8023
totalVar	1.8474	1.0000	1.359	8.155	1.000	0.8155

* Contrib equals contribution in %

**Number of Distinct Categories (truncated signal-to-noise-ratio) = 7

Interpretação dos resultados

Componentes da variação

O Rstudio também calcula uma coluna de componentes de variação (**Varcomp**), que são os valores base para calcular %Gage R&R usando-se o método ANOVA.

As estimativas para um estudo Gage R&R são obtidas através do cálculo dos componentes de variação de cada fonte de variação.

A repetibilidade e a reprodutibilidade contém os componentes operador e a interação operador e banco. Juntos eles compõem a variação do sistema de medição.

Por que usar componentes de variação ?

Utilizamos componentes de variação para conhecer a contribuição de cada fonte de variação na variação total.

Operators:	2	Parts:	10			
Measurements:	2	Total:	40			
<hr/>						
Anova Table - crossed Design						
Df Sum Sq Mean Sq F value Pr(>F)						
Operator	1	0.00	0.000	0.006 0.938		
Part	9	65.07	7.230	177.429 <2e-16 ***		
Operator:Part	9	0.70	0.078	1.915 0.109		
Residuals	20	0.81	0.041			
<hr/>						
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						
<hr/>						
Gage R&R						
	VarComp	VarCompContrib	StDev	StudyVar	StudyVarContrib	P/T Ratio
totalRR	0.0594	0.0321	0.244	1.462	0.179	0.1462
repeatability	0.0407	0.0221	0.202	1.211	0.149	0.1211
reproducibility	0.0186	0.0101	0.137	0.819	0.100	0.0819
Operator	0.0000	0.0000	0.000	0.000	0.000	0.0000
Operator:Part	0.0186	0.0101	0.137	0.819	0.100	0.0819
Part to Part	1.7881	0.9679	1.337	8.023	0.984	0.8023
totalVar	1.8474	1.0000	1.359	8.155	1.000	0.8155
<hr/>						

* Contrib equals Contribution in %						
**Number of Distinct Categories (truncated signal-to-noise-ratio) = 7						

Interpretação dos resultados

As tabelas Gage R&R mostram como a variabilidade total é dividida entre as seguintes fontes :

- Precisão (**Total Gage R&R**), que pode ser dividida em:
 - Repetibilidade (**Repeatability**), que é a variabilidade nas medidas obtida quando a mesma peça é medida múltiplas vezes pelo mesmo operador;
 - Reprodutibilidade (**Reproducibility**) (que pode ainda ser dividida nos componentes operador e operador e peça), que é a variabilidade nas medidas obtida quando a mesma peça é medida por operadores diferentes.
- Processo produtivo (**Part-to-Part**), que é a variabilidade nas medidas de peças diferentes

Em condições ideais, muito pouco da variabilidade deveria dever-se à repetibilidade e reproducibilidade. Ao invés disso, a diferença entre as peças deveria responder pela maior parte da variabilidade.

Operators:	2	Parts:	10			
Measurements:	2	Total:	40			
<hr/>						
Anova Table - crossed Design						
Df	Sum Sq	Mean Sq	F value	Pr(>F)		
Operator	1	0.00	0.000	0.006 0.938		
Part	9	65.07	7.230	177.429 <2e-16 ***		
Operator:Part	9	0.70	0.078	1.915 0.109		
Residuals	20	0.81	0.041			
<hr/>						
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						
<hr/>						
Gage R&R						
	VarComp	VarCompContrib	stdev	studyVar	studyVarContrib	P/T Ratio
totalRR	0.0594	0.0321	0.244	1.462	0.179	0.1462
repeatability	0.0407	0.0221	0.202	1.211	0.149	0.1211
reproducibility	0.0186	0.0101	0.137	0.819	0.100	0.0819
Operator	0.0000	0.0000	0.000	0.000	0.000	0.0000
Operator:Part	0.0186	0.0101	0.137	0.819	0.100	0.0819
Part to Part	1.7881	0.9679	1.337	8.023	0.984	0.8023
totalVar	1.8474	1.0000	1.359	8.155	1.000	0.8155
<hr/>						
* Contrib equals contribution in %						
**Number of Distinct Categories (truncated signal-to-noise-ratio) = 7						

Interpretação dos resultados

Porcentagem de contribuição

%Contribution, baseada nos componentes de variação, é calculada dividindo-se cada valor da coluna VarComp pela variação total (Total Variation) e, em seguida, multiplica-se o resultado por 100.

Por exemplo, para calcular a %Contribution de Part-to-Part, divida VarComp de Part-to-Part por Total Variation e multiplique o resultado por 100.

$$(1,7881 / 1,8474) \times 100 = 96,79$$

O valor Part-to-Part é de 96,79. Isso significa que 96,79% da variação na medição é devida à diferença entre as peças. Isso é muito bom. É desejável que a %Contribution de Part-to-Part seja elevada porque isso significa que o sistema é capaz de distinguir as peças umas das outras.

Uso de variação vs. desvio padrão

Uma vez que %Contribution baseia-se na variância, sua soma será 100%. É possível também mostrar colunas com percentuais baseados no desvio padrão (ou raiz quadrada da variância) de cada termo. Estas colunas, rotuladas de percentual da variação em estudo (%StudyVar), percentual de tolerância (%Tolerance) e percentual de variação do processo (%Process), não somarão 100%. Uma das vantagens de se usar o desvio padrão é que pode-se comparar a variação com a tolerância, uma vez que ambas estarão na mesma unidade de medida.

Operators:	2	Parts:	10	
Measurements:	2	Total:	40	
<hr/>				
ANOVA Table - crossed Design				
Df	Sum Sq	Mean Sq	F value	Pr(>F)
Operator	1	0.00	0.000	0.006 0.938
Part	9	65.07	7.230	177.429 <2e-16 ***
Operator:Part	9	0.70	0.078	1.915 0.109
Residuals	20	0.81	0.041	
<hr/>				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
<hr/>				
Gage R&R				
totalRR	0.0594	0.0321	0.244	1.462 0.179 0.1462
repeatability	0.0407	0.0221	0.202	1.211 0.149 0.1211
reproducibility	0.0186	0.0101	0.137	0.819 0.100 0.0819
operator	0.0000	0.0000	0.000	0.000 0.000 0.0000
Operator:Part	0.0186	0.0101	0.137	0.819 0.100 0.0819
Part to Part	1.7881	0.9679	1.337	8.023 0.984 0.8023
totalVar	1.8474	1.0000	1.359	8.155 1.000 0.8155

* Contrib equals contribution in %

**Number of Distinct Categories (truncated signal-to-noise-ratio) = 7

Interpretação dos resultados

Estudo de variação percentual

Usamos **%StudyVar** quando estamos interessados em comparar a variação do sistema de medição com a variação total.

%StudyVar é calculada dividindo-se cada valor em **StudyVar** por **Total Variation** e multiplicando-se o resultado por 100.

StudyVar é calculado como aproximadamente 6 vezes o desvio padrão para cada fonte.

Por que usar 6 ?

Normalmente, a variação do processo é definida como 6σ (desvio padrão como estimativa de σ). Em um estudo de R&R, 6 é padrão porque, quando os dados encontram-se distribuídos normalmente, 99% deles irá cair dentro de desvios padrão de 6.

Operators:	2	Parts:	10			
Measurements:	2	Total:	40			
<hr/>						
Anova Table - crossed Design						
Df Sum Sq Mean Sq F value Pr(>F)						
Operator	1	0.00	0.000	0.006 0.938		
Part	9	65.07	7.230	177.429 <2e-16 ***		
Operator:Part	9	0.70	0.078	1.915 0.109		
Residuals	20	0.81	0.041			
<hr/>						
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						
<hr/>						
Gage R&R						
	VarComp	VarCompContrib	stdev	studyVar	studyVarContrib	P/T Ratio
totalRR	0.0594	0.0321	0.244	1.462	0.179	0.1462
repeatability	0.0407	0.0221	0.202	1.211	0.149	0.1211
reproducibility	0.0186	0.0101	0.137	0.819	0.100	0.0819
Operator	0.0000	0.0000	0.000	0.000	0.000	0.0000
Operator:Part	0.0186	0.0101	0.137	0.819	0.100	0.0819
Part to Part	1.7881	0.9679	1.337	8.023	0.984	0.8023
totalVar	1.8474	1.0000	1.359	8.155	1.000	0.8155

* Contrib equals contribution in %

**Number of Distinct Categories (truncated signal-to-noise-ratio) = 7

Interpretação dos resultados

Percentual de tolerância

Frequentemente, deseja-se comparar a variação do sistema de medição com a especificação para determinar a proporção da especificação usada pela variação no sistema de medição.

Ao digitar a especificação, o MINITAB calculará **%Tolerance**, que compara a variação do sistema de medição com as especificações do cliente.

A variação do sistema de medição ($6 \times SD$) é dividida pela especificação (faixa de medição permitida pelo cliente). A proporção resultante é multiplicada por 100 e reportada como **%Tolerance**.

Aqui, **%Tolerance** para o R&R é **14,62%**, o que não é muito bom.

Qual medida usar ?

A escolha entre **%Tolerance** e **%StudyVar** dependerá da aplicação do sistema de medição.

- Se o sistema de medição for usado para melhoria de processos (redução a variação processo produtivo), o **%StudyVar** oferecerá uma estimativa melhor da precisão da medição;
- Se o sistema de medição for usado para avaliar as peças em relação às especificações, **%Tolerance** será uma medida mais apropriada.

Operators:	2	Parts:	10			
Measurements:	2	Total:	40			
<hr/>						
Anova Table - crossed Design						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Operator	1	0.00	0.000	0.006	0.938	
Part	9	65.07	7.230	177.429	<2e-16 ***	
Operator:Part	9	0.70	0.078	1.915	0.109	
Residuals	20	0.81	0.041			
<hr/>						
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						
<hr/>						
Gage R&R						
	VarComp	VarCompContrib	stdev	studyVar	studyVarContrib	P/T Ratio
totalRR	0.0594	0.0321	0.244	1.462	0.179	0.1462
repeatability	0.0407	0.0221	0.202	1.211	0.149	0.1211
reproducibility	0.0186	0.0101	0.137	0.819	0.100	0.0819
Operator	0.0000	0.0000	0.000	0.000	0.000	0.0000
Operator:Part	0.0186	0.0101	0.137	0.819	0.100	0.0819
Part to Part	1.7881	0.9679	1.337	8.023	0.984	0.8023
totalVar	1.8474	1.0000	1.359	8.155	1.000	0.8155
<hr/>						

* Contrib equals contribution in %

**Number of Distinct Categories (truncated signal-to-noise-ratio) = 7

Interpretação dos resultados

Total de R&R do instrumento (Total Gage R&R)

Seu maior interesse encontra-se na fileira Total Gage R&R dessa tabela.

Neste caso, não importa se está considerando **%StudyVar** ou **%Tolerance** : a contribuição do sistema de medição para a variação geral no estudo é superior a 10% (Total Gage R&R) em ambos os casos.

Lembre-se que a diferença entre **%Tolerance** e **%StudyVar** é o divisor. Uma vez que a faixa de especificação (10) é mais ampla que a variação do processo (7), os percentuais para **%Tolerance** serão mais baixos.

NESTE ESTUDO O SISTEMA DE MEDIÇÃO DEVE SER MELHORADO OU SUBSTITUÍDO PARA GARANTIR A EFICIÊNCIA DO PROCESSO PRODUTIVO !!!

Operators:	2	Parts:	10
Measurements:	2	Total:	40
<hr/>			
Anova Table - crossed Design			
Df	Sum Sq	Mean Sq	F value
Operator	1	0.00	0.006
Part	9	65.07	7.230
Operator:Part	9	0.70	0.078
Residuals	20	0.81	0.041
<hr/>			
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			
<hr/>			
Gage R&R			
totalRR	0.0594	0.0321	0.244
repeatability	0.0407	0.0221	0.202
reproducibility	0.0186	0.0101	0.137
Operator	0.0000	0.0000	0.000
Operator:Part	0.0186	0.0101	0.137
Part to Part	1.7881	0.9679	1.337
totalVar	1.8474	1.0000	1.359
<hr/>			
* Contrib equals Contribution in %			
**Number of Distinct Categories (truncated signal-to-noise-ratio) = 7			

Interpretação dos resultados

Número de categorias distintas

O número de categorias distintas informa quantos grupos distintos de peças o sistema de medição é capaz de distinguir.

O número de categorias distintas é calculado por :

$$\frac{S_{\text{peça}}}{S_{\text{sistema de medição}}} \times 1,414$$

e, em seguida, arredondando para o número inteiro seguinte.

Número de categorias	Significa que...
< 2	O sistema não consegue discriminar as peças umas das outras.
2	As peças podem ser divididas em grupos altos e baixos. Isso equivale aos dados de atributos.
≥ 5	O sistema é aceitável e consegue distinguir as peças umas das outras

Aqui, o número de categorias é 7, o que indica que o sistema é capaz de distinguir as peças uma das outras.

```

Operators:      2      Parts:   10
Measurements:  2      Total:   40
-----
Anova Table - crossed Design
  Df Sum Sq Mean Sq F value Pr(>F)
Operator       1    0.00   0.000   0.006  0.938
Part          9   65.07   7.230 177.429 <2e-16 ***
Operator:Part  9    0.70   0.078   1.915  0.109
Residuals     20   0.81   0.041
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
-----
Gage R&R
  VarComp VarCompContrib stdev studyVar studyVarContrib P/T Ratio
totalRR        0.0594      0.0321 0.244   1.462      0.179  0.1462
repeatability  0.0407      0.0221 0.202   1.211      0.149  0.1211
reproducibility 0.0186      0.0101 0.137   0.819      0.100  0.0819
  Operator      0.0000      0.0000 0.000   0.000      0.000  0.0000
  Operator:Part 0.0186      0.0101 0.137   0.819      0.100  0.0819
Part to Part   1.7881      0.9679 1.337   8.023      0.984  0.8023
totalVar       1.8474      1.0000 1.359   8.155      1.000  0.8155
---
* Contrib equals Contribution in %
**Number of Distinct Categories (truncated signal-to-noise-ratio) = 7

```

Interpretação dos resultados

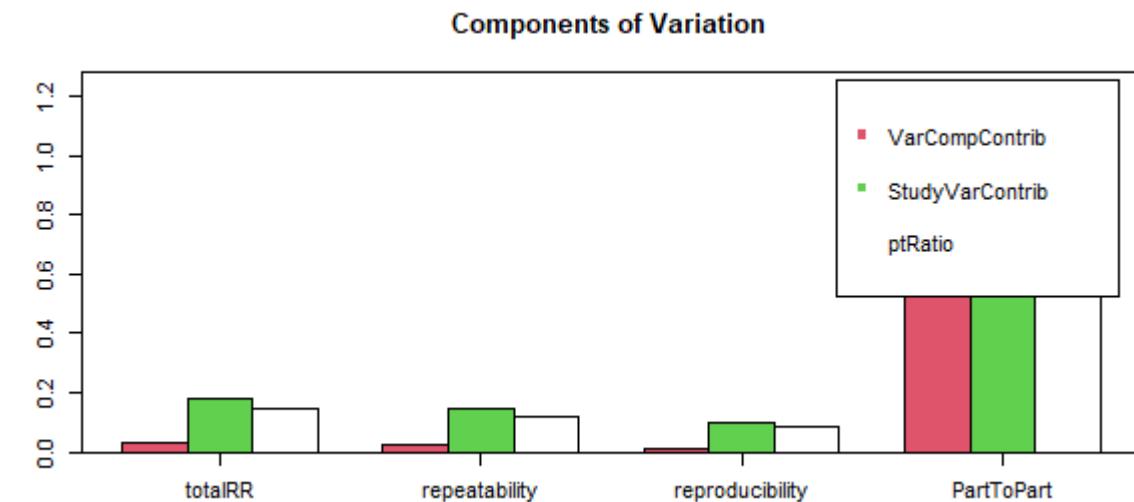
Componentes de variação

O gráfico **Components of variation** é uma representação visual da tabela Gage R&R na saída da janela Session.

Cada agrupamento de barras representa uma fonte de variação. Por default, cada agrupamento terá duas barras, que correspondem a %Contribution e %StudyVar. Se for acrescentada uma tolerância e/ou sigma histórico, serão acrescentadas barras para %Tolerance e/ou %Process.

Em um bom sistema de medição, o maior componente de variação é a variação do processo produtivo. Se, ao invés disso, obtiver grandes volumes de variação atribuídos a Gage R&R (repetibilidade e/ou reproduzibilidade), será preciso tomar uma ação corretiva.

Para os dados correspondentes ao torque dos bancos, a responsabilidade pela maior parte da variação é da diferença entre as peças, conforme pode ser observado no gráfico abaixo.



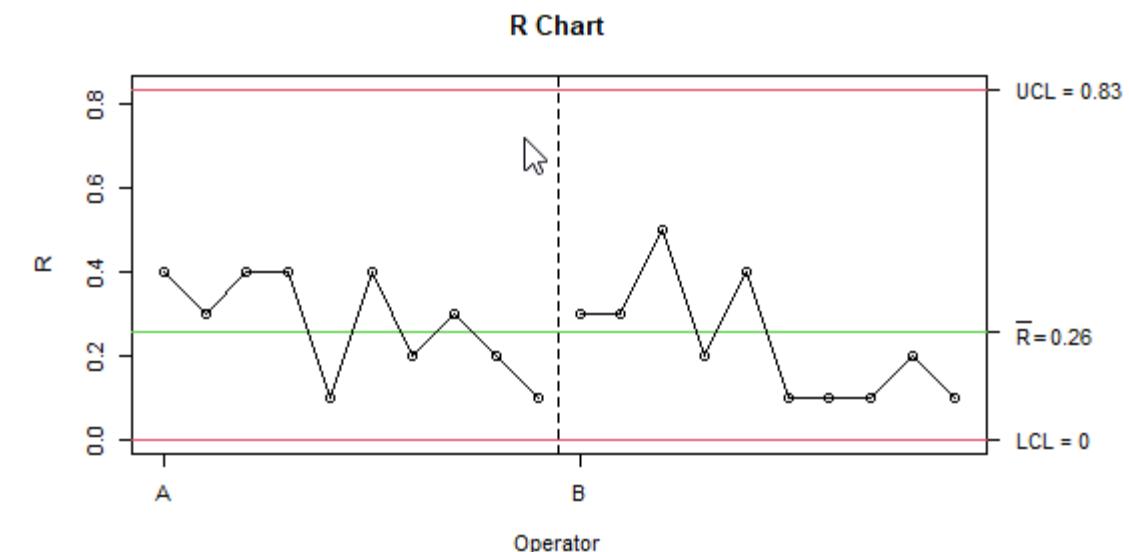
Interpretação dos resultados

Carta R

A carta R consiste no seguinte :

- Os pontos marcados na carta representam, para cada operador, a diferença entre a maior e a menor medição de cada peça. Se as medições forem as mesmas, a amplitude será igual a zero. Uma vez que os pontos são distribuídos por operador, é possível comparar a consistência de cada operador.
- A linha central, que é a média total do processo (média para todas as faixas de subgrupos)
- Os limites de controle representam a variação esperada entre as medidas do mesmo operador. Estes limites são calculados utilizando-se as variações dentro dos subgrupos (operadores).

Se qualquer um dos pontos na carta ultrapassar o limite superior de controle (UCL), isso significa que o operador está tendo problemas ao medir as peças. O valor UCL leva em consideração o número de medições feitas por um operador e a variabilidade entre as peças. Se os operadores estiverem tomando medições de forma consistente, essas faixas deveriam ser pequenas em relação aos dados e os pontos deveriam ficar dentro da faixa de controle.



Interpretação dos resultados

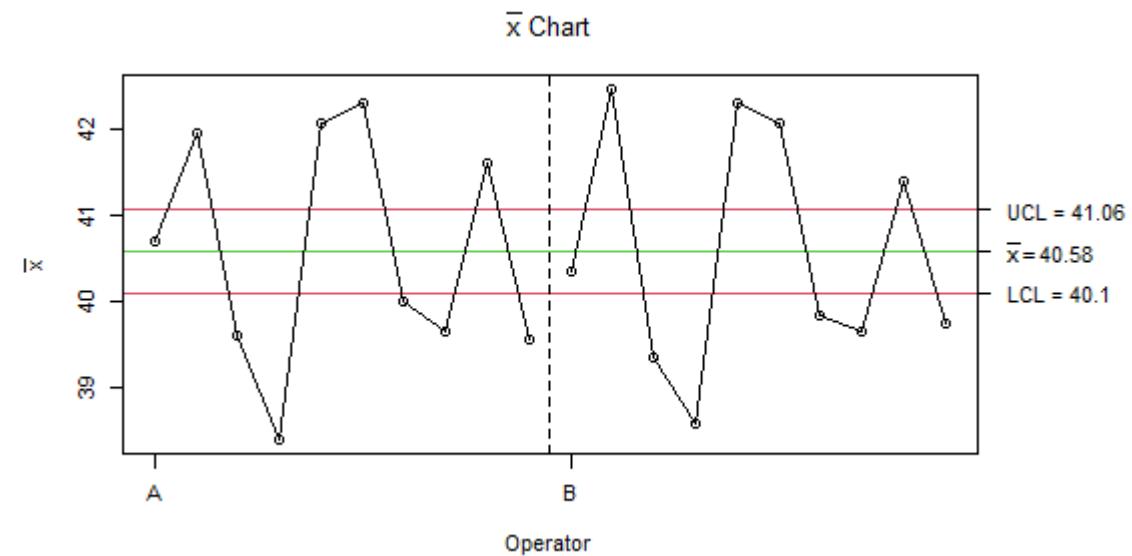
Carta X

A carta X compara a variação do processo produtivo com a repetibilidade. A carta X consiste no seguinte:

- Cada ponto representa a média das medições das peças feitas por um operador, em uma amostra
- A linha central, que é a média geral de todas as peças medidas por todos os operadores
- Os limites de controle (UCL e LCL), baseiam-se na variabilidade entre as peças.

Uma vez que as peças escolhidas para um Gage R&R devem representar toda a faixa de peças possíveis, em condições ideais essa carta deveria mostrar falta de controle. Existe falta de controle quando muitos pontos estão acima do limite superior de controle e/ou abaixo do limite inferior de controle.

Em relação aos dados das peças, não existem pontos além dos limites de controle, o que indica que o sistema de medição não é adequado, como já visto anteriormente.



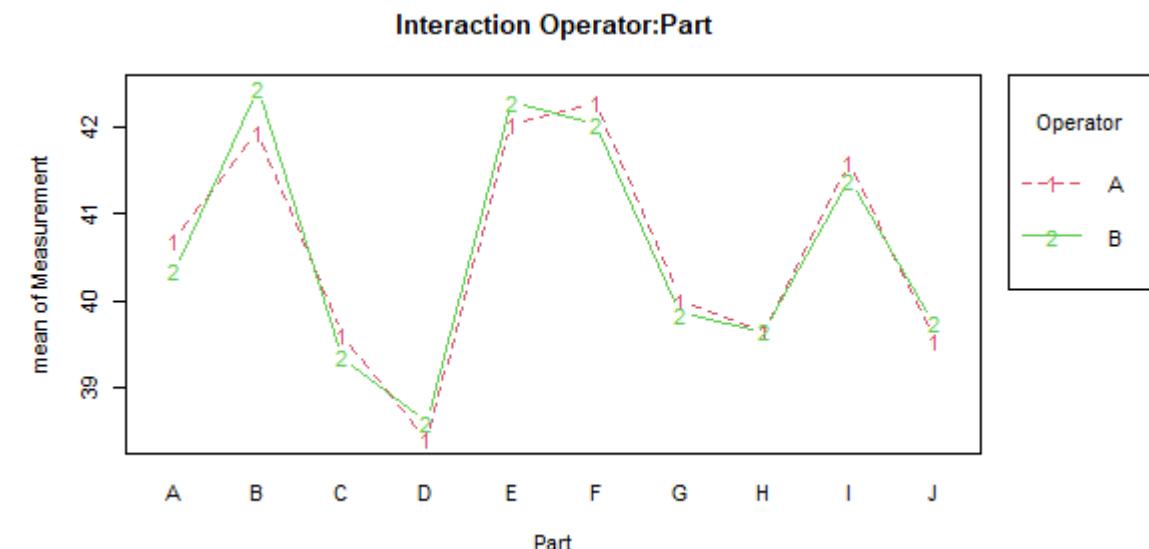
Interpretação dos resultados

Interação entre operador e peça

O gráfico de interação entre operador e peça mostra a média das medidas realizadas por cada operador em cada peça. Para cada operador, uma linha conecta as médias das medidas das peças. Em condições ideais, as linhas deveriam seguir o mesmo padrão e as médias das peças deveriam variar de forma tal que as diferenças entre as peças fique clara.

Padrão	Significa que...
As linhas são virtualmente idênticas	Os operadores que estão medindo as peças são a mesma pessoa
Uma linha é consistentemente mais alta ou mais baixa do que as outras	Determinado operador está medindo as peças de forma consistentemente mais alta ou mais baixa do que os demais
As linhas não são paralelas ou se cruzam	A capacidade do operador de medir uma peça depende de qual peça está sendo medida (uma interação entre Operador e Peça)

Aqui, as linhas seguem uma a outra bem de perto e as diferenças entre as peças estão claras.



Interpretação dos resultados

Medições por operador

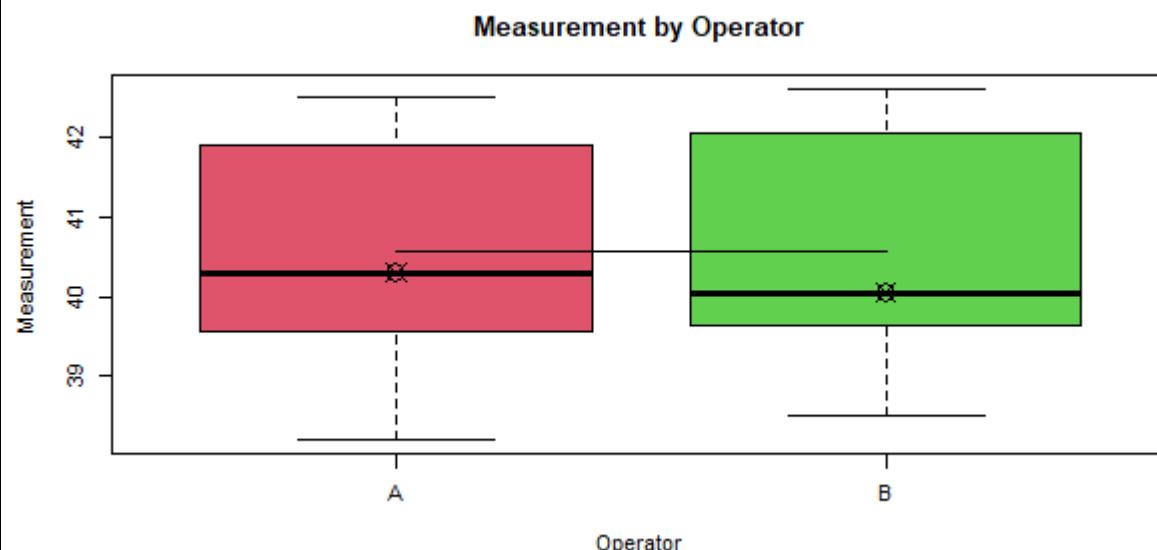
O gráfico de efeitos principais **By Operador** pode ajudar a determinar se as medições e a variabilidade são consistentes em relação ao operador.

O gráfico **By Operador** mostra todas as medições do estudo dispostas por operador. Os pontos representam as medições; os símbolos de círculos com cruzes representam as médias. A linha vermelha conecta as medidas das médias para cada operador.

Se a linha vermelha...	então...
estiver paralela ao eixo de x	os operadores estão medindo as peças de forma similar
não estiver paralela ao eixo de x	os operadores estão medindo as peças de forma diferente

Também é possível determinar se a variabilidade geral na medição das peças é a mesma usando-se este gráfico. A dispersão nas medições é similar? Ou, um operador varia mais que os outros?

Aqui, os operadores parecem estar medindo as peças de forma consistente, com aproximadamente a mesma variabilidade.



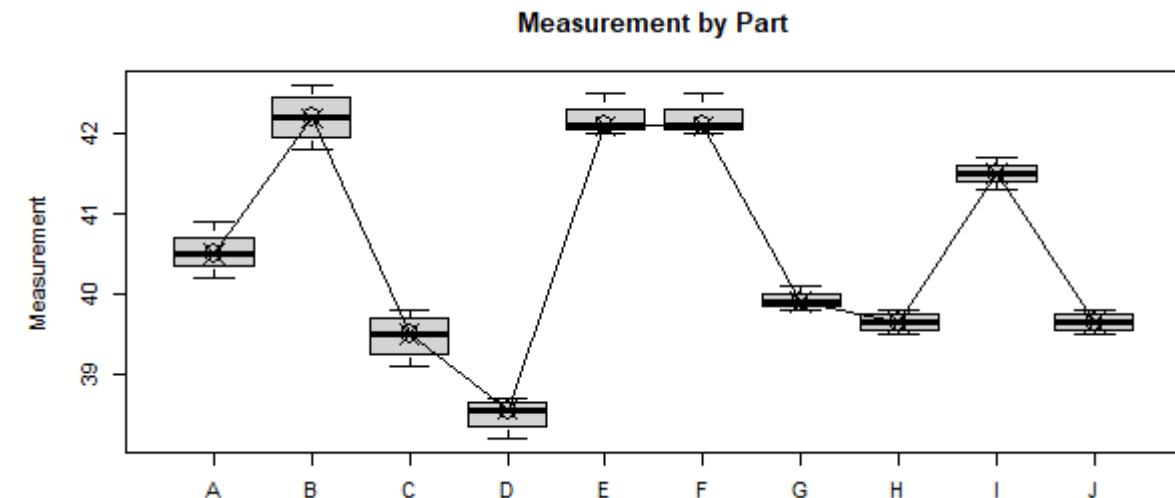
Interpretação dos resultados

Medições por peça

O gráfico **By Part** mostra todas as medições realizadas no estudo, dispostas por peça. As medições são representadas por pontos; as médias, pelo símbolo de círculo com cruz.

Em condições ideais :

- Múltiplas medições para cada peça apresentam pouca variação (os pontos para uma peça estarão bem próximos entre si)
- As médias irão variar de tal forma que as diferenças entre as peças fiquem claras



Considerações Finais

Conclusões práticas

Nesse exemplo, o sistema de medição contribui de forma negativa para a variação geral.

A variação devida ao sistema de medição, ou seja, o **%StudyVar** ou o **%Tolerance**, é superior a 10%. De acordo com as diretrizes **AIAG**, este é um sistema aceitável.

Com base nestes resultados, não podemos nos sentir plenamente seguros, pois o sistema de medição contribui de alguma forma, para a variação total do processo.

Conclusões estatísticas

Utilize a tabela e os gráficos R&R para determinar a variação do sistema de medição. Utilize as diretrizes a seguir para a tabela R&R de instrumento.

%Tolerance; %StudyVar	%Contribution	O sistema é...
10% ou menos	1% ou menos	Ideal
10% - 20%	1% - 4%	Aceitável
20% - 30%	4% - 9%	Marginal
30% ou mais	10% ou mais	Ruim

Os padrões de gráfico abaixo mostram baixa variação no sistema de medição :

- pequena faixa média na carta Rbar
- limites de controle estreitos na carta Xbar e muitos pontos fora de controle
- linhas sobrepostas no gráfico de interação operador e peça
- uma linha horizontal no gráfico por operador
- medições individuais muito similares para cada peça em relação a todos os operadores no gráfico por peça e diferenças óbvias entre as peças



Nível Sigma

Nível: Intermediário

Nível Sigma - Variáveis

Nível Sigma – Variáveis

Calculando Nível Sigma – Variáveis

```
81 #Análise de Capabilidade
82 library("SixSigma")
83 piece<- read_excel("ES01246.xlsx")
84 piece.values<- piece$Altura
85 ss.ca.z(piece.values, LSL = 31.1, USL = 33.1)
86 ss.ca.cp(piece.values, LSL = 31.1, USL = 33.1, ci = TRUE)
87 ss.ca.cpk(piece.values, LSL = 31.1, USL = 33.1, ci = TRUE)
88 ss.study.ca(piece.values, LSL = 31.1, USL = 33.1, Target = 32.1)
89
90 #outra forma
91 library("qualityTools")
92 library("qcc")
93 altura<- qcc(piece.values, type = "xbar.one", nsigmas = 3, plot = FALSE)
94 process.capability(altura, spec.limits = c(31.1, 33.1))
```

Aqui, carregamos a biblioteca SixSigma, carregamos os dados na variável piece, pegamos apenas a Altura na variável piece.values, geramos os índices Z-Bench, Cp, Cpk; sempre definindo os limites inferior e superior de spec., pedindo também o intervalo de confiança.

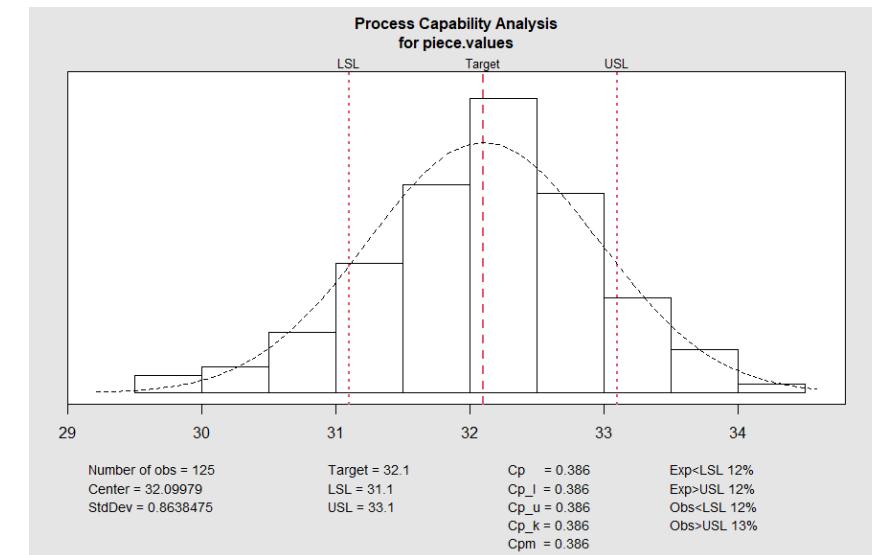
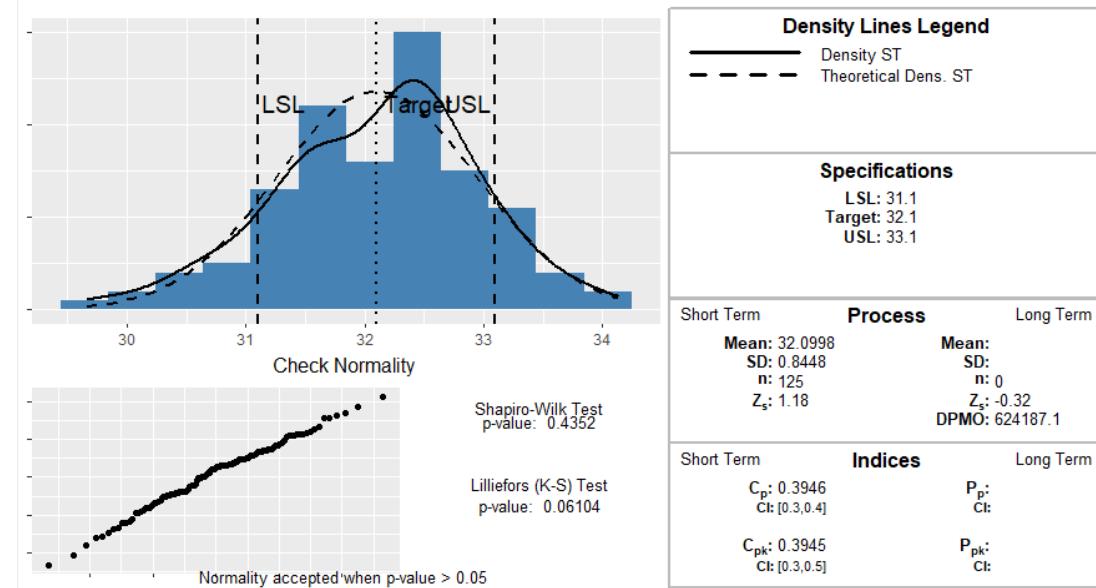
Por fim, pedimos o estudo de capacidade colocando o valor target, para gerar o Cpm

A resposta sai em forma de texto e gráfico.

Na sequência mostro outra biblioteca que também faz o gráfico da capacidade

```
> ss.ca.z(piece.values, LSL = 31.1, USL = 33.1)
[1] 1.183504
> ss.ca.cp(piece.values, LSL = 31.1, USL = 33.1, ci = TRUE)
[1] 0.3455032 0.4435925
> ss.ca.cpk(piece.values, LSL = 31.1, USL = 33.1, ci = TRUE)
[1] 0.3181774 0.4708250
> ss.study.ca(piece.values, LSL = 31.1, USL = 33.1, Target = 32.1)
```

Six Sigma Capability Analysis Study



Interpretação dos resultados

Histograma

O RStudio fornece um histograma com a distribuição dos dados, onde fica visível que o processo segue uma distribuição normal. Neste histograma estão dispostos os limites de controle (inferior e superior). Fica também visível, a existência de pontos fora dos limites de controle, o que demonstra instabilidade.

Caixa de Informações

O RStudio fornece caixas de informações, onde estão dispostos :

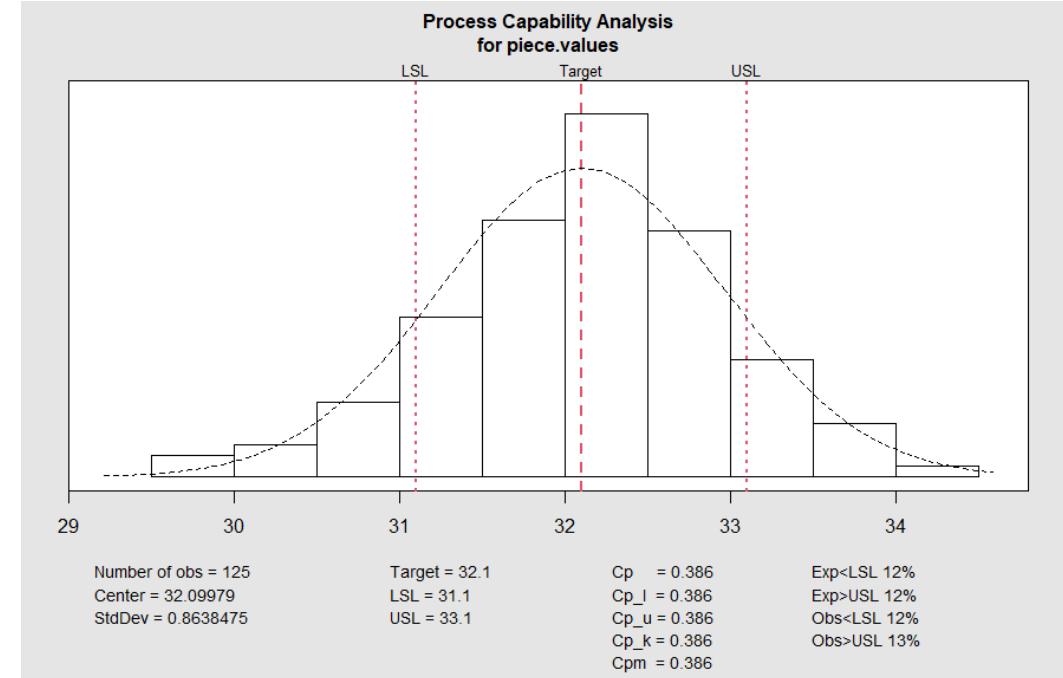
- Média e desvio padrão do processo
- PPM observado (real) e PPM esperado à curto e à longo prazo mas em percentual

Capabilidade do processo

Além das informações citadas acima, o MINITAB fornece :

- Z-Bench de 1,18, ou seja, o nível sigma
- Cp - Coeficiente de capabilidade do processo (0,39)
- Cpk - Coeficiente de capabilidade preliminar do processo (0,39)

**Lembrando, cada 0,33 = 1 sigma



Process Capability Analysis

```
call:  
process.capability(object = altura, spec.limits = c(31.1, 33.1))
```

Métrica	Valor
Number of obs	125
Center	32.1
StdDev	0.8638
Target	32.1
LSL	31.1
USL	33.1

Capability indices:

	value	2.5%	97.5%
Cp	0.3859	0.3379	0.4338
Cp_l	0.3858	0.3223	0.4493
Cp_u	0.3860	0.3225	0.4494
Cp_k	0.3858	0.3102	0.4614
Cpm	0.3859	0.3381	0.4336

Métrica	Valor
Exp<LSL	12%
Exp>USL	12%
Obs<LSL	12%
Obs>USL	13%

R7

Ferramentas do Planejamento da Qualidade

Nível: Intermediário

Gráficos de Pareto

O que é um gráfico de Pareto ?

Um gráfico de Pareto é um gráfico de barras. Existem dois tipos básicos de gráficos de Pareto :

- Regular – Os defeitos são ordenados do mais comum ao menos comum;
- Ponderado – Os defeitos são ordenados por considerações tais como custo, severidade ou detectabilidade. Um gráfico de Pareto ponderado proporciona informações similares às obtidas por meio de uma Análise do Modo e Efeito da Falha (Failure Modes Effect Analysis – FMEA)

Quando usar um gráfico de Pareto ?

Utilize um gráfico de Pareto quando tiver contagens de defeitos. Os dados devem ser estruturados de uma das seguintes maneiras :

- Dados brutos em uma única coluna – Cada observação é uma ocorrência de um tipo de defeito em particular. Pode-se também incluir uma variável de agrupamento como, por exemplo, turnos, em uma segunda coluna. Isso lhe permitirá criar gráficos de Pareto em separados para cada turno.

- Dados sumarizados em uma tabela de duas colunas – Uma coluna contém os tipos de defeitos e a outra, as contagens para cada tipo de defeito. Pode-se também incluir um fator de peso como, por exemplo, custo em uma terceira coluna. Isso lhe permitirá ordenar os defeitos no gráfico de Pareto por custo.

Por que usar um gráfico de Pareto ?

Utilize um gráfico de Pareto para priorizar problemas e esforços em focos de melhoria em áreas onde podem ser obtidos os maiores ganhos. Se for preciso levar em consideração questões tais como custo, severidade e detectabilidade, utilize um gráfico de Pareto ponderado.

Um gráfico de Pareto pode ajudá-lo a responder perguntas como, por exemplo :

- Quais defeitos ocorrem com maior frequência em nossos produtos ?
- Quais defeitos estão gerando as maiores despesas ?
- Certos defeitos são mais comuns durante certos turnos ?

Criando um gráfico de Pareto

É necessário determinar que tipos de defeitos em bancos ocorrem com maior frequência em nossa linha de montagem.

Todos os defeitos são apontados no último posto de trabalho da linha de montagem, o que irá gerar o indicador IPPM. Identificar os defeitos mais comuns irá ajudar a focar os esforços para melhoria deste indicador.

Crie um gráfico de Pareto para identificar os defeitos que ocorrem com maior frequência na linha de montagem.

Gráfico de Pareto

```
101 #Ferramentas da Qualidade
102 #Pareto
103 ippm<- read_excel("IPPM.xlsx")
104 ippm.values<- ippm$Quantidade
105 help("pareto.chart")
106 help("paretoChart")
107 names(ippm.values)<- ippm$Defeito
108 pareto.chart(ippm$Quantidade)
109 #outro melhor
110 paretoChart(ippm.values, showTable = TRUE)
111
```

Temos sempre mais de uma forma de fazer a mesma coisa. Nesse caso apresentaremos duas.

Carregamos os dados IPPM, colocamos na variável ippm.values.

Aqui uma coisa interessante: daremos nomes às barras, com a função names()

Geramos um tipo de pareto e depois outro tipo, note a diferença

Ferramentas do Planejamento da Qualidade

Interpretação dos resultados

O gráfico de Pareto mostra a classificação de defeitos, do mais comum ao menos comum. Aqui, 29,7% do total de defeitos é devido a revestimento com rugas, vincos e dobras. A segunda maior categoria, revestimentos com pontas de linha, responde por 24,8% do total de defeitos. Juntos, revestimento com rugas, vincos e dobras e revestimento com pontas de linha, respondem por quase 55% do total de defeitos.

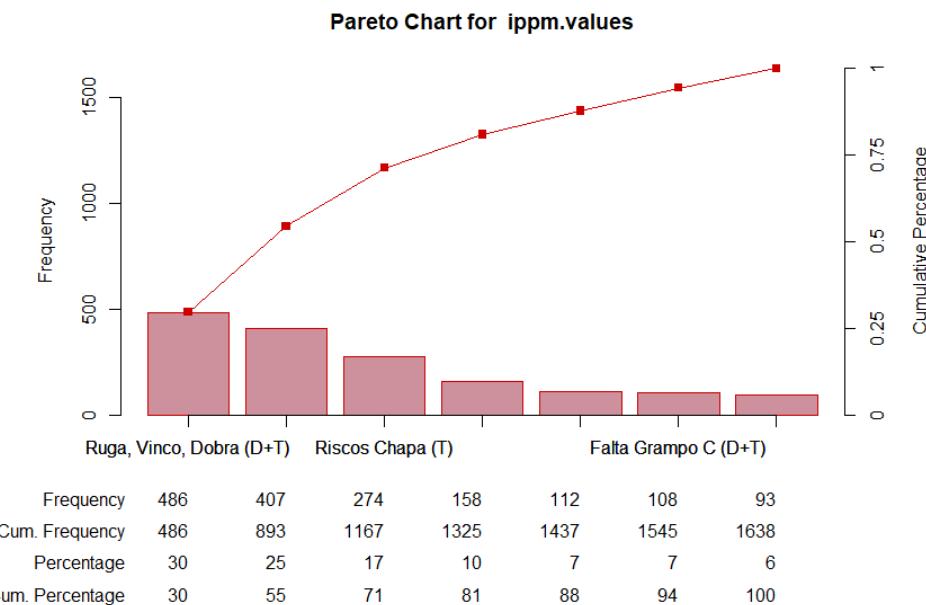
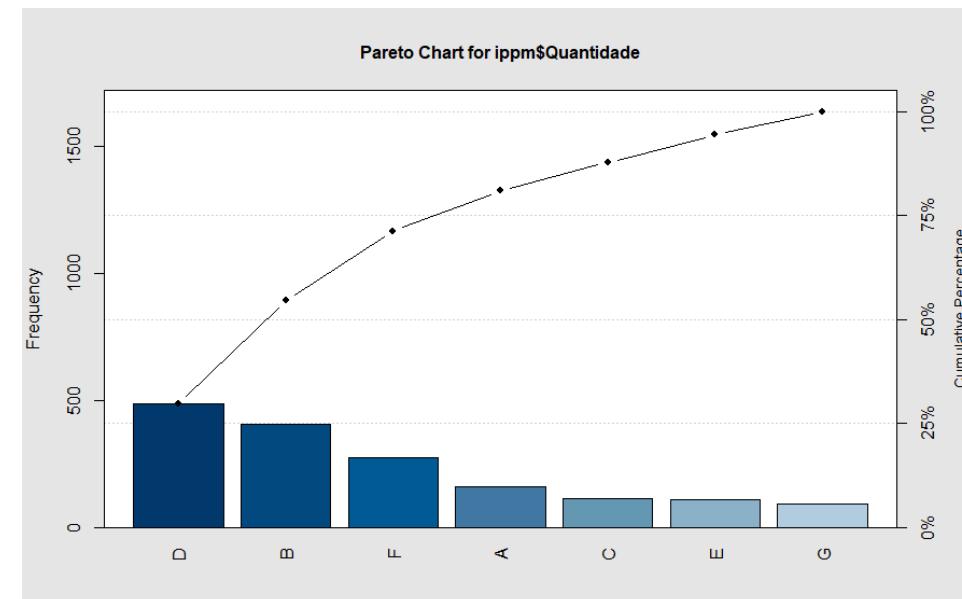
O MINITAB exibe uma categoria, rotulada de outros (**Others**), quando os defeitos restantes representam menos de 5% da contagem total de defeitos. É possível alterar o tamanho da categoria Others para qualquer valor que desejar (Basta alterar o valor digitado no campo **Combine defects after the first __% into one**).

Conclusões

Aqui, 29,7% dos defeitos correspondem a revestimentos com rugas, vincos e dobras. Portanto, é desejável que os esforços de melhoria sejam focados na identificação das razões para tantos revestimentos com este tipo de defeito e fazer um brainstorming para obter possíveis soluções.

O defeito revestimento com rugas, vincos e dobras, continua sendo um dos mais importantes se ao invés de ocorrência, passarmos a analisar o custo ?

Para responder a esta pergunta, pode-se desenvolver um gráfico de Pareto ponderado por custo (Basta alterar o item **Frequencies in de Quantidade para Custo**).



Considerações Finais

Conclusões práticas

O gráfico de Pareto inicial indicou que revestimento com rugas, vincos e dobras e revestimento com pontas de linha, são os defeitos que ocorrem mais comumente.

Antes de focar os esforços na resolução destes problemas, decidiu-se criar um gráfico de Pareto ponderado para ver quais defeitos contribuíam mais para os custos da qualidade.

O gráfico de Pareto ponderado mostrou que é possível obter maiores ganhos focando na solução dos problemas de riscos na chapa, almofada mau colada e botão de comando enroscando (estamos falando de ganhos financeiros, neste caso).

Conclusões estatísticas

Algumas vezes, um gráfico de Pareto de contagens pode não proporcionar todas as informações necessárias para se tomar uma decisão.

Por exemplo, um defeito que ocorre com baixa frequência pode apresentar um custo muito elevado. Em um gráfico de Pareto de contagens, este defeito não será um daqueles considerados dentre os poucos que são vitais. Se você atribuir custos aos defeitos poderá criar um gráfico de Pareto ponderado que irá ajudar a identificar defeitos menos comuns, mas de custo mais elevado. Os defeitos também podem ser ponderados por gravidade ou outros fatores.

Diagrama de Causa e Efeito

O que é um diagrama de causa e efeito ?

Um diagrama de causa e efeito, também conhecido como diagrama fishbone (espinha de peixe) ou Ishikawa, organiza idéias de um brainstorming sobre as causas potenciais de um problema. O diagrama de causa e efeito pode ajudá-lo a visualizar as relações entre um problema e suas possíveis causas associadas.

Quando usar um diagrama de causa e efeito ?

Utilize um diagrama de causa e efeito quando tiver até 6 (seis) colunas contendo causas potenciais.

Por que usar um diagrama de causa e efeito ?

O diagrama de causa e efeito é utilizado para separar problemas de grande porte em partes fáceis de administrar.

Por exemplo, você pode usar um diagrama de causa e efeito para :

- separar as fontes potenciais de problemas em um processo de montagem ?
- fazer um brainstorming e levantar os riscos potenciais em uma fábrica ?

Criando um diagrama de causa e efeito

O propósito do diagrama de causa e efeito é organizar idéias sobre causas potenciais dos problemas de qualidade.

Liste as causas potenciais para grupo de causas em colunas distintas. Pode-se usar até seis grupos de causas.

Por default, o MINITAB usa as seguintes categorias para rotular as categorias do diagrama : Pessoal (**Personnel**), Máquinas (**Machines**), Materiais (**Materials**), Métodos (**Methods**), Medidas (**Measurements**) e Ambiente (**Environment**).

Pode-se alterar os rótulos default do MINITAB para outros que se adequem às necessidades individuais.

Gráfico de Pareto

```
112 #ishikawa
113 help("cause.and.effect")
114 ishikawa<- read_excel("FISHBONE.xlsx")
115 cause.and.effect(cause = list(Medição=ishikawa$Medição,
116                               Material=ishikawa$`Matéria Prima`,
117                               Mão_de_Obra=ishikawa$`Mão de Obra`,
118                               Meio_Ambiente=ishikawa$`Meio Ambiente`,
119                               Método=ishikawa$Método,
120                               Máquina=ishikawa$Máquina),
121                               effect = "IPPM Alto")
```

Neste caso, carregamos o arquivo e precisamos dar nome às colunas e atribuir valores

No que no Rstudio não para dar quebra de linha basta ir para a próxima linha com Enter, ele continua lendo sem problema

O código é dividido em 2 argumentos

- Cause: é uma lista onde damos, por exemplo, o nome medição e atribuímos a coluna Medição da variável ishikawa, que é nosso arquivo
- Effect: que pode ser uma variável, uma sétima coluna no arquivo, um valor digitado, como quiser.

Feito isso o gráfico é gerado e apresentado no próximo slide.

Interpretação dos resultados

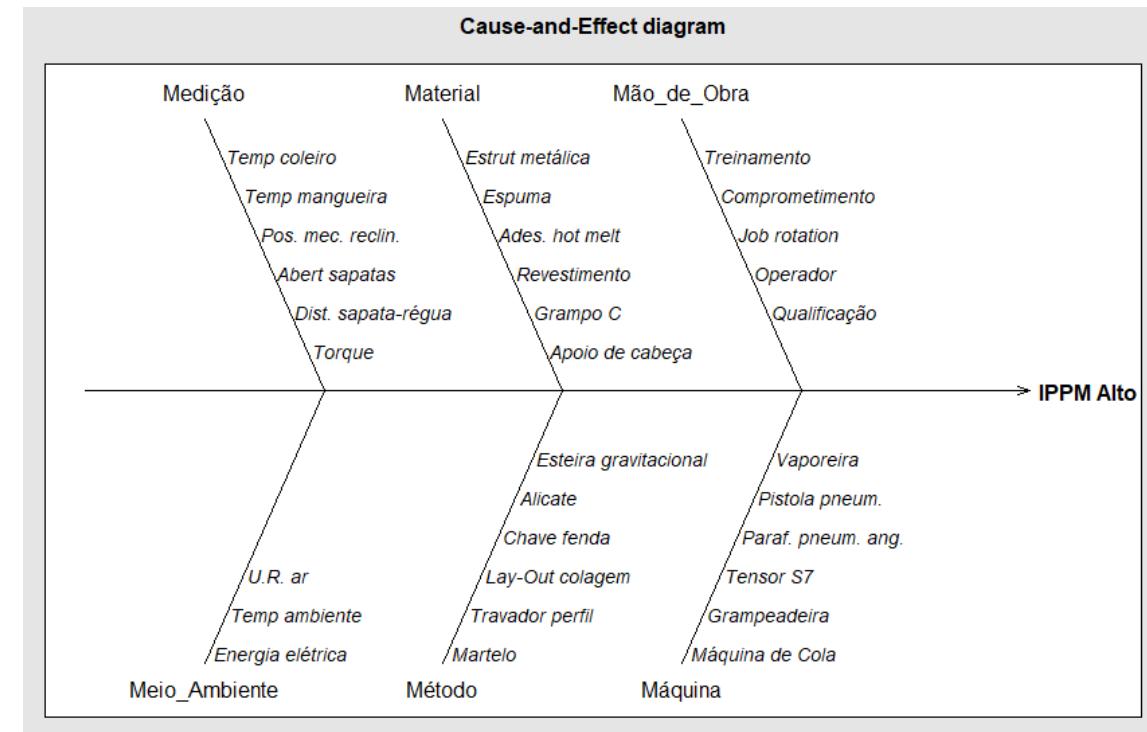
Cada categoria do diagrama de causa e efeito exige uma coluna de causas.

Criando sub-níveis

Pode-se listar causas secundárias criando outro diagrama.

Por exemplo, digamos que o efeito do problema seja **Personnel** (Mão de Obra) e as causas relativas a este efeito são **Treinamento, Comprometimento, Job Rotation, Falha operador e Qualificação**.

Estas causas potenciais se tornaram as categorias do novo diagrama de causa e efeito para o efeito **Personnel** (Mão de Obra). No código do RStudio cada causa deve ser carregada em uma coluna distinta.



Considerações Finais

Conclusões práticas

Após organizar as causas potenciais, a equipe deve classificá-las com base em sua relevância. Em seguida, equipes menores irão focar em cada categoria para resolver a maioria das causas do problema.

Outras considerações

- Utilize o conhecimento do setor da garantia da qualidade sobre o processo para criar um diagrama de causa e efeito eficiente
- Assegure-se de ter uma compreensão completa do processo; faça isso solicitando brainstormings de todos os membros da equipe
- Priorize uma categoria do diagrama de causa e efeito de cada vez. Desdobre esta categoria em vários outros diagramas

R 8

Testes de Hipóteses para Dados Variáveis

Nível: Avançado

Teste de hipóteses

O que é teste de hipóteses?

Um teste de hipótese utiliza dados de uma amostra para testar uma hipótese sobre a população da qual a amostra foi obtida.

Quando usar teste de hipóteses?

Use testes de hipóteses quando tiver dados de uma amostra e desejar fazer inferências sobre uma ou mais populações.

Por que usar um teste de hipóteses?

Testes de hipóteses podem ajudá-lo a responder a perguntas tais como :

- O processo está corretamente centrado?
- O produto de um fornecedor é melhor do que o produto de outro?
- Existem diferenças entre grupos de tratamento no experimento?

Por exemplo :

- O estoque de papel tem, em média, 21,6 cm de largura ?
- O combustível do fornecedor A queima de forma mais limpa do que o combustível do fornecedor B ?
- Qual a fórmula de refrigerante preferida pelos clientes?

Testando a hipótese nula

Precisamos determinar se a média do processo de conformação do grampo C difere de maneira significativa da dimensão nominal de 20,25 milímetros. Em termos estatísticos, a média do processo é também chamada de *média de população*, ou μ (mu).

Hipótese estatística

Existem 2 possibilidades : ou média é igual a 20,25 ou não é. Essas alternativas podem ser declaradas como 2 hipóteses :

- A *hipótese nula* (H_0) : μ igual 20,25 mm
- A *hipótese alternativa* (H_a) : μ não é igual 20,25 mm

Uma vez que não é viável medir cada caixa de grampo C na população, nunca poderemos saber com certeza qual hipótese é a correta. Contudo, um teste de hipótese apropriado pode ajudar-nos a dar um palpite com um certo grau de certeza.

PARA ANÁLISE DESSES DADOS, O TESTE APROPRIADO É O TESTE t DE UMA AMOSTRA (1 Sample-t).

A) Teste t de 1 amostra

```
131 #Testes de Hipóteses
132 #teste t para 1 amostra
133 grampo<- read_excel("GRAMPO C.xlsx")
134 help(t.test)
135 t.test(grampo$`Dim Interna`, mu=20.25)
```

Esse é fácil, apenas carregue os dados e, com uma linha de Código já fazemos o teste.

Queremos comparar se a média dos dados são ou não iguais a 20,25 (por isso o argumento `mu=20.25`)

Como resposta temos:

```
One Sample t-test

data: grampo$`Dim Interna`
t = -138.7, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 20.25
95 percent confidence interval:
 20.02493 20.03127
sample estimates:
mean of x
 20.0281
```

Interpretando os resultados

O resultado do teste está ao lado:

Lembrando :

- A *hipótese nula* (H_0): μ igual 20,25 mm
- A *hipótese alternativa* (H_a): μ não é igual 20,25 mm

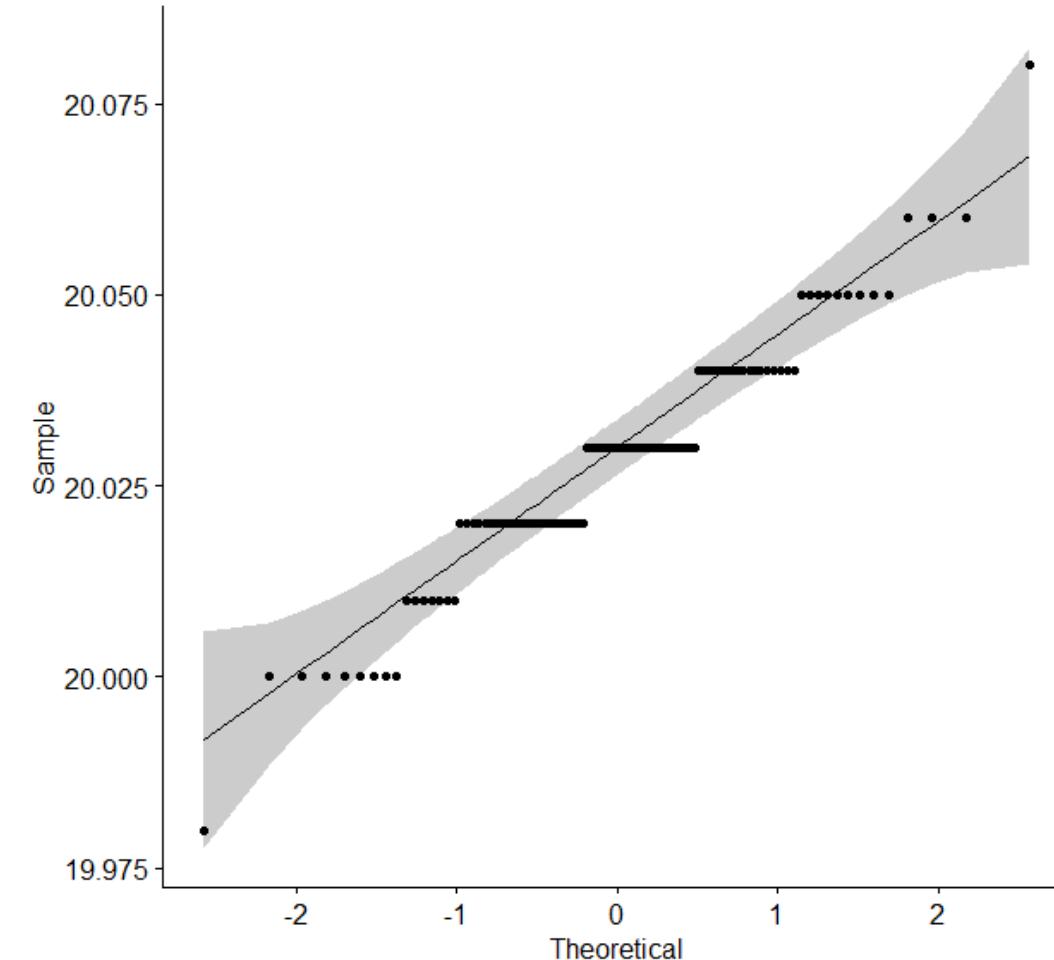
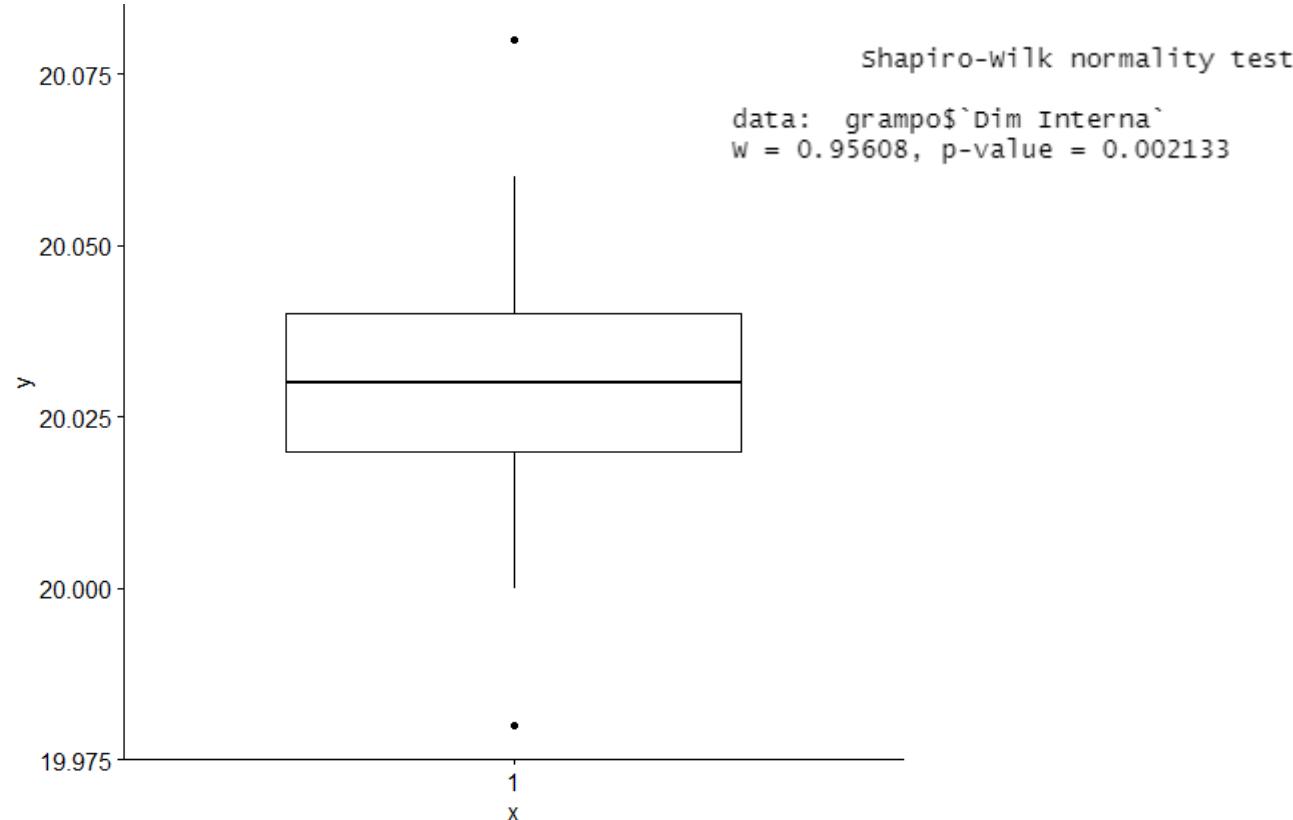
- ✓ Se o **valor-p** é menor ou igual a α , rejeitamos H_0
- ✓ Se o **valor-p** é maior do que α , não rejeitamos H_0

Como poderemos verificar, o **valor-p** é 0,000. Como este valor é menor que α (0,05), rejeitamos a hipótese nula, ou seja, a média do processo de conformação do grampo C quanto à característica dimensão interna, **não é igual** à média prevista na especificação (dimensão nominal).

O desvio padrão do estudo é de **0,0160** e o intervalo de confiança está entre **20,0249 e 20,0313** (nível de confiabilidade de 95%).

Bônus – Pacote ggppubr para análise de dados

```
137 #Bônus para análise dos dados!
138 install.packages("ggpubr")
139 library("ggpubr")
140 ggboxplot(grampo$`Dim Interna`) # outro tipo de boxplot para aprender
141 shapiro.test(grampo$`Dim Interna`) #Outro tipo de teste de normalidade
142 ggqqplot(grampo$`Dim Interna`) # outro tipo de teste de probabilidade
143
```



B) Teste t de 2 amostras

```
144 #Teste t para duas amostras
145 install.packages("rstatix")
146 library("rstatix")
147 grampo2<- read_excel("GRAMPO C2.xlsx")
148 #opção 1: desconsiderando variâncias iguais
149 t.test(grampo2$`DI Probel`, grampo2$`DI Grampofix`, var.equal = FALSE)
150 #opção 2: considerando variâncias iguais
151 t.test(grampo2$`DI Probel`, grampo2$`DI Grampofix`, var.equal = TRUE)
```

No teste t para duas amostras, simplesmente adicionamos o segundo termo à função t.test

No primeiro caso, não forçamos a condição de que as variâncias são iguais. Assim o resultado é:

```
welch Two Sample t-test

data: grampo2$`DI Probel` and grampo2$`DI Grampofix`
t = -180.02, df = 165.88, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.5444733 -0.5326601
sample estimates:
mean of x mean of y
20.02810 20.56667
```

Na outra já dizemos para considerar variâncias iguais, e:

```
Two Sample t-test

data: grampo2$`DI Probel` and grampo2$`DI Grampofix`
t = -179.54, df = 196, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.5443997 -0.5325700
sample estimates:
mean of x mean of y
20.02818 20.56667
```

Interpretando os resultados

O resultado do teste está abaixo :

Lembrando :

- A *hipótese nula* (H_0) : $\mu_{\text{DI Probel}} = \mu_{\text{DI Grampofix}}$
- A *hipótese alternativa* (H_a) : $\mu_{\text{DI Probel}} \neq \mu_{\text{DI Grampofix}}$

Como poderemos verificar, o **valor-p** é 0,000. Como este valor é menor que **a** (0,05), rejeitamos a hipótese nula, ou seja, a média do processo do fornecedor A (Probel) **não é igual** à média do processo do fornecedor B (Grampofix).

O desvio padrão do estudo é de **0,0160** para Probel e **0,0252** para Grampofix. O intervalo de confiança está entre **-0,544 e -0,533** (nível de confiabilidade de 95%).

C) Teste para Variâncias Iguais

No teste t de 2 amostras apresentado nos slides anteriores, assumimos que a variância do processo da Probel é igual a variância do processo da Grampofix. O RStudio permite a realização de um teste que valida se os processos possuem ou não a mesma variância, é o *Teste para Variâncias Iguais*.

```
153 #Teste para iguais variâncias  
154 #para tirar a dúvida acima, de qual usar:  
155 var.test(grampo2$`DI Probel`, grampo2$`DI Grampofix`)  
156
```

Nesse comando é só trocar o t por var.

```
F test to compare two variances  
  
data: grampo2$`DI Probel` and grampo2$`DI Grampofix`  
F = 0.40457, num df = 99, denom df = 98, p-value = 1.034e-05  
alternative hypothesis: true ratio of variances is not equal to 1  
95 percent confidence interval:  
 0.2718754 0.6017711  
sample estimates:  
ratio of variances  
 0.4045653
```

Interpretando os resultados

Existem 2 testes, nos quais podemos analisar o valor-p:

F-Test (dados normais) e Levene's Test (dados não necessariamente normais). Não testamos a normalidade dos dados, porém o valor-p é 0,000, ou seja, os processos não possuem variâncias iguais como assumido no teste anterior. O correto seria realizar novamente o teste t de 2 amostras, assumindo que a variância dos processos não são iguais.

D) Teste T pareado

O que é um teste t pareado ?

Um teste t pareado ajuda a determinar se a diferença média entre observações pareadas é significativa. Estatisticamente, ele é equivalente a executar um teste t nas diferenças de cada observação pareadas. Um teste t pareado também pode ser usado para avaliar se a diferença média é igual a um valor específico.

Observações pareadas se relacionam de alguma forma. Os exemplos incluem :

- ✓ Pesos de indivíduos registrados antes e depois de um programa de exercícios
- ✓ Medidas da mesma peça tomadas com 2 dispositivos de medição diferentes

Quando usar um teste t pareado ?

Use um teste t pareado em uma amostra aleatória de observações pareadas. Os dados devem ser contínuos (quantitativos).

Por que usar um teste t pareado ?

Um teste t pareado pode ajudar a responder questões tais como :

- ✓ Um novo tratamento provoca uma diferença no produto?
- ✓ Dois instrumentos medem a mesma coisa, com a mesma precisão?

Executando o teste t pareado

Estamos com alguns problemas dimensionais com a bucha do apoio de cabeça do Gol (VW) e as inspeções são realizadas pelo departamento de metrologia e pelos inspetores da qualidade, com instrumentos diferentes. Gostaríamos de saber se ambos os instrumentos são confiáveis, ou seja, se realizam as medições com as mesmas faixas de incerteza.

Uma vez que os dados são pareados (cada bucha do apoio de cabeça sendo dimensionada por ambos os dispositivos), use **Paired t** para testar as seguintes hipóteses:

- ✓ H_0 : a diferença média entre observações pareadas na população é zero
- ✓ H_a : a diferença média entre observações pareadas na população não é zero

Você ainda pode criar dotplots e boxplots para visualizar os dados.

Use um nível de confiança padrão de 95% para o teste.

t pareado

Sem segredo, carregamos o dataset e pedimos o teste t, incluindo um argumento a mais: paired = TRUE

```
157 #Teste t pareado
158 pareado<- read_excel("APCAB.xlsx")
159 t.test(pareado$Instr_A, pareado$Instr_B, paired = TRUE)
160
```

E a resposta:

```
Paired t-test

data: pareado$Instr_A and pareado$Instr_B
t = -2.8132, df = 19, p-value = 0.0111
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.025287867 -0.003712133
sample estimates:
mean of the differences
-0.0145
```

Interpretando os resultados

As dimensões médias da bucha do apoio de cabeça para o instrumento B é 12,8190 milímetros e para o instrumento A é 12,8045 milímetros. A diferença média é de 0,0145 milímetros.

O intervalo de confiança de 95%, para a diferença média são está entre 0,003712 e 0,025288.

Estatística t e valor-p

O teste dá uma estatística t de 2,81, o qual está associado com um valor-p de 0,011. Assim, podemos rejeitar a hipótese nula no nível α de 0,05 e concluir que a diferença média entre observações na população, não é zero (H_a), ou seja, podemos afirmar que na dimensão de uma mesma peça, executada com equipamentos diferentes, os valores encontrados serão diferentes.

Pense e responda rápido :

**NOSSO SISTEMA DE MEDIÇÃO É CONFIÁVEL?
PODEMOS CONTINUAR NOSSOS ESTUDOS?**

E) Análise de poder

O poder é a capacidade de um teste detectar um defeito quando este existir.

Pode ser entendido ainda, como a probabilidade de que a rejeição da hipótese nula está correta, dado que a hipótese nula é falsa.

Podemos usar uma análise de poder para determinar quanto poder tem um teste existente ou para ajudar a projetar um novo teste, de modo que este tenha o poder adequado.

Quando usar uma análise de poder ?

Use uma análise de poder quando estiver planejando um experimento ou após executar um teste de hipótese. Não são necessários quaisquer dados. Será preciso estimar o desvio padrão (exceto para testes de proporção, aplicados para dados qualitativos, os quais veremos mais a frente).

Por que usar uma análise de poder ?

Uma análise de poder pode ajudar a responder a questões, como por exemplo :

- ✓ O tamanho da amostra é grande o bastante?
- ✓ Qual o tamanho da diferença que é capaz de detectar com o teste?
- ✓ Podemos confiar nos resultados não significativos de um teste?

Determinando o poder

Gostaríamos de saber qual o poder do teste realizado para validar as alterações realizadas na formulação da espuma do banco dianteiro.

Temos 12 amostras, o desvio padrão é de 2,43 e a diferença entre a formulação atual e a proposta é 3,12 (teste t de 2 amostras). *Desejamos obter um poder de 95%.*

Roteiro para execução do estudo :

```
161 #Poder amostral
162 install.packages("pwr")
163 library("pwr")
164 help("pwr.t.test")
165 d = 3.12/2.43 #Effect Size (Cohen's d) Differences / STDEV
166 pwr.t.test(n=12, d=d, type = "two.sample")
```

Note que, diferentemente do minitab, ao invés de colocar a diferença e o desvio padrão em separado, deve-se fazer uma conta antes, o D de Cohen. E a resposta vem no próximo slide:

Interpretando os resultados

```
Two-sample t test power calculation
```

```
n = 12  
d = 1.283951  
sig.level = 0.05  
power = 0.8520314  
alternative = two.sided
```

```
NOTE: n is number in *each* group
```

Como podemos observar, com este número de amostras obtemos um poder de 85,2%, quando desejávamos 95%.

Aqui devemos tomar uma decisão :

- ✓ Realizar o estudo nestas condições e assumir um erro de aproximadamente 15% (muitas vezes este é o único caminho, devido a dificuldade de se conseguir amostras); ou
- ✓ Aumentar o número de amostras e alcançar um poder de 95% e aí sim realizar o estudo e tirar as conclusões desejadas.

Definindo o novo número de amostras

Agora vamos inverter. Queremos saber quantas amostras preciso ter para um poder de 95%.

```
167 pwr.t.test(d=d, power=.95, type = "two.sample")
```

Uma linha de código resolve o problema.

```
Two-sample t test power calculation
```

```
n = 16.79015  
d = 1.283951  
sig.level = 0.05  
power = 0.95  
alternative = two.sided
```

```
NOTE: n is number in *each* group
```

Como não Podemos amostrar 16,7 coisas, sempre arredondamos para cima, ou seja 17.

R 9

Testes de Hipóteses para Proporções (Atributos)

Nível: Avançado

A) Teste de Uma Proporção

O que é um teste de uma proporção?

Um teste de uma proporção ajuda a determinar se a proporção de uma população é diferente de um valor específico (proporção específica).

Quando usar um teste de uma proporção ?

Use um teste de uma proporção para avaliar dados de proporção para uma única amostra.

Por que usar um teste de uma proporção ?

Um teste de uma proporção pode ajudar a responder questões tais como :

- ✓ A proporção da população difere de 0,5 ?
- ✓ Uma proporção é maior ou menor que um critério ?

Por exemplo :

- ✓ Um programa de inteligência artificial é capaz de responder Sim / Não a questões com uma precisão maior do que 50% ?
- ✓ O índice de defeito de prendedores plásticos está abaixo do máximo aceitável ?

Executando o teste de uma proporção

Queremos divulgar para nossos clientes em potencial, que nosso índice de retrabalho é menor em relação ao índice de retrabalho de nosso principal concorrente (XYZ), porém gostaríamos de ter certeza antes de tomar esta atitude.

O índice mensal de retrabalho do nosso concorrente XYZ é de 4,0% em média.

Mensalmente produzimos em média 71.668 bancos (17.917 kits), sendo que destes 2.772 bancos passam pelo posto de retrabalho por algum motivo.

Usaremos um teste de uma proporção para determinar se essa diferença é significativa. As hipóteses para o teste são :

- ✓ **H₀** : Nossa índice de retrabalho é igual ao índice de retrabalho de nosso concorrente (4%);
- ✓ **H_a** : Nossa índice de retrabalho não é igual ao índice de retrabalho de nosso concorrente (4%).

Usaremos um nível de confiança de 95%.

Executando o teste de uma proporção

- 1 O número de peças é **71668** (*Produção*)
- 2 A quantidade de retrabalhos é **2772**
- 3 Queremos testar se essa proporção é igual a 4%

```
169 #Teste de proporções
170 prop.test(2772, 71668, .04)
```

Simples como isso. A resposta é:

```
1-sample proportions test with continuity correction
data: 2772 out of 71668, null probability 0.04
X-squared = 3.2257, df = 1, p-value = 0.07249
alternative hypothesis: true p is not equal to 0.04
95 percent confidence interval:
 0.03728431 0.04012209
sample estimates:
    p 
0.03867835
```

Para não usar a correção é só acrescentar o termo `Correction = FALSE`.

Interpretação dos resultados

O α usado nesse teste é 0,05.

Os resultados sugerem que o nosso índice de retrabalho (**Sample p = 0,038678**) é mais baixo que o índice de retrabalho do nosso concorrente XYZ que é de 0,04, porém :

- O intervalo de confiança de 95% (0,037278 a 0,040116) inclui 0,04 (4%, retrabalho do concorrente XYZ);
- O valor-p (0,070) é maior que α (0,05).

Sendo assim, não rejeitamos a hipótese nula, ou seja, não podemos dizer para nossos clientes potenciais que o nosso índice de retrabalho é menor que o índice de retrabalho do concorrente XYZ.

B) Teste de Duas Proporções

O que é um teste de duas proporções ?

Um teste de duas proporções ajuda a determinar se a proporção de uma população é diferente em relação à proporção de uma segunda população.

Quando usar um teste de duas proporções ?

Use um teste de duas proporções para avaliar dados de proporção para duas amostras diferentes.

Por que usar um teste de duas proporções ?

Um teste de duas proporções pode ajudar a responder questões tais como :

✓ Uma proporção é maior ou menor que uma outra proporção ?

Por exemplo :

✓ O índice de grampos defeituosos do fornecedor A é maior, menor ou igual ao índice de grampos defeituosos do fornecedor B ?

✓ O índice de bancos rejeitados no primeiro turno é igual ao índice de bancos rejeitados no segundo turno ?

Executando o teste de duas proporções

Nosso fornecedor de revestimentos, está sob regime de embarque controlado (CS1), devido a excesso de peças com ponta de linha durante o 1º semestre de 2005. O fornecedor implementou uma série de melhorias no 2º semestre de 2005 e vem afirmando que com estas melhorias o índice de revestimentos com pontas de linha foi reduzido e por isso quer que a Lear o retire do regime CS1.

1º semestre : 106.411 bancos produzidos e detectados 3.718 revestimentos com ponta de linha

2º semestre : 85.419 bancos produzidos e detectados 3.215 revestimentos com ponta de linha

Usaremos um teste de duas proporções para determinar se essa diferença é significativa para o nosso processo.

As hipóteses para o teste são :

✓ **H₀** : O índice de revestimentos com pontas de linha no 1º semestre é igual ao índice de revestimentos com pontas de linha no 2º semestre;

✓ **H_a** : O índice de revestimentos com pontas de linha no 1º semestre não é igual ao índice de revestimentos com pontas de linha no 2º semestre

Usaremos um nível de confiança de 95%.

Executando o teste de duas proporções

```
171 prop.test(x = c(3718, 3215), n = c(106411, 85419))
```

Simples assim.

Não sei se já havia comentando antes, mas, em R quando queremos colocar 2 itens em um argumento, usamos a função `c()` que combina os valores.

A resposta é:

```
2-sample test for equality of proportions

data: c(3718, 3215) out of c(106411, 85419)
X-squared = 9.824, df = 1, p-value = 0.001723
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.004395625 -0.001000374
sample estimates:
 prop 1   prop 2
0.034940 0.037638
```

Interpretação dos resultados

O α usado nesse teste é 0,05.

Os resultados sugerem que o índice de revestimentos com pontas de linha no 1º semestre (**Sample p = 0,034940**) é mais baixo que o índice de revestimentos com pontas de linha no 2º semestre (**Sample p = 0,037638**).

- O intervalo de confiança de 95% (0,00101093 a 0,0438507) inclui a estimativa da diferença (0,002698);
- O valor-p (0,002) é menor que α (0,05).

Sendo assim, rejeitamos a hipótese nula, ou seja, podemos dizer que o índice de revestimentos com ponta de linha no **1º semestre é diferente** do índice de revestimentos com pontas de linha no **2º semestre**. Se observarmos atentamente os dados, veremos que houve uma **elevação** no índice de revestimentos não conformes, o que inviabiliza retirar o regime de embarque controlado do fornecedor no momento.



10

Regressão

Nível: Avançado

A) Correlação

O que é correlação ?

O coeficiente de correlação da amostra, "r", mede o grau de associação linear entre duas variáveis (o grau em que uma variável varia em relação a outra).

Quando usar correlação ?

Use correlação quando possuir dados de duas variáveis contínuas e desejar determinar se há uma relação linear entre elas. A correlação não dirá se as variáveis estão relacionadas de maneira não linear.

Alguns estatísticos acreditam que a correlação não deveria ser usada se uma variável for uma resposta dependente da outra.

Por que usar correlação ?

A correlação pode ajudar a responder questões tais como :

- Duas variáveis se relacionam de maneira linear ?
- Qual é a força da relação ?

Por exemplo :

- Há uma relação entre a temperatura e a viscosidade de um óleo para cozinha ?
- Quão forte é a relação entre exposição a ultravioleta e a resistência de uma trama de náilon para tendas ?

Visualizando os dados

```
175 #Regressão
176 #Scatterplot e correlação
177 hotmelt<- read_excel("HOTMELT.xlsx")
178 head(hotmelt)
179 scatter.smooth(hotmelt$Temperatura, hotmelt$Tempo)
180 #outro jeito
181 plot(hotmelt$Temperatura, hotmelt$Tempo)
182 #Cálculo de correlação
183 help(cor.test)
184 cor(hotmelt$Temperatura, hotmelt$Tempo)
185 cor.test(hotmelt$Temperatura, hotmelt$Tempo, method = "pearson")
```

Aproveitei para já colocar todos os comandos próximos.

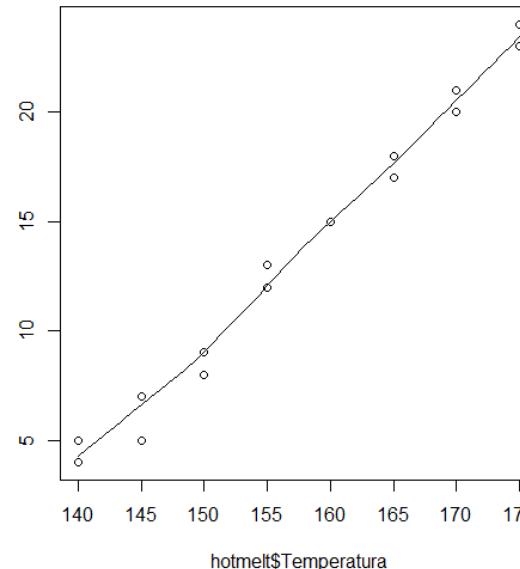
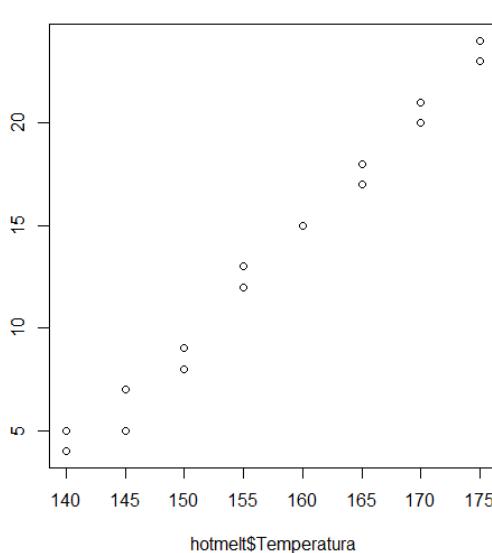
1. Carregamos os dados
2. Visualizamos as primeiras linhas
3. Montamos um tipo de scatterplot
4. Depois outro tipo
5. Pedimos o valor da correlação
6. Geramos um relatório de correlação

Interpretação dos resultados

O gráfico de **Temperatura** versus **Tempo** indica :

Há uma forte relação entre as características (temperatura da máquina de cola e tempo de secagem e fixação do Hot Melt). Quando os valores de **Tempo** mudam, os valores de **Temperatura** também mudam;

- ✓ Os dados seguem em uma linha razoavelmente reta, o que sugere que a relação é linear;
- ✓ Valores elevados da característica **Temperatura** estão associados com valores elevados da característica **Tempo**, o que indica que a relação é positiva.



Calculando a correlação

O coeficiente de correlação de Pearson é calculado para determinar a força da associação linear entre as características avaliadas, neste caso, **Temperatura** e **Tempo**.

Pearson's product-moment correlation

```
data: hotmelt$Temperatura and hotmelt$Tempo
t = 32.792, df = 14, p-value = 1.22e-14
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9809995 0.9978217
sample estimates:
cor
0.9935532
```

Kendall's rank correlation tau

```
data: hotmelt$Temperatura and hotmelt$Tempo
z = 5.0482, p-value = 4.459e-07
alternative hypothesis: true tau is not equal to 0
sample estimates:
tau
0.965546
```

Spearman's rank correlation rho

```
data: hotmelt$Temperatura and hotmelt$Tempo
S = 5.0304, p-value = 3.188e-14
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.9926024
```

Interpretação dos resultados

```
> cor(hotmelt$Temperatura, hotmelt$Tempo)
[1] 0.9935532
```

Pearson correlation (r) :

O valor de r sempre estará entre -1 e 1 :

- 1 indica uma correlação positiva perfeita
- 0 indica ausência de correlação
- -1 indica uma correlação negativa perfeita

Valor-p :

O valor-p testa as seguintes hipóteses :

- H_0 : o coeficiente de correlação (ρ ou rho) para a relação entre as populações é igual a zero;
- H_a : ρ não é igual a zero.

Conclusão :

O coeficiente de correlação (0,994) indica que há uma forte associação linear positiva entre as características

Temperatura e Tempo.

Além disso, o valor-p (0,000) é menor que a (0,05); portanto, pode-se rejeitar a hipótese nula de que não existe associação linear.

Considerações estatísticas

A correlação quantifica o grau de associação linear entre duas variáveis.

Uma forte correlação não implica uma relação de causa e efeito. Por exemplo, uma forte correlação entre 2 variáveis pode ser devida à influência de uma terceira variável que não foi analisada.

Um coeficiente de correlação próximo de zero não significa necessariamente ausência de associação, apenas que a associação não é linear. Sempre grafe os dados de modo que possa identificar relações não lineares quando estas estiverem presentes.

Alguns estatísticos argumentam que a correlação não deveria ser usada se uma variável for resposta dependente da outra.

Correlações pressupõem que os valores das duas variáveis variam livremente. Não é permitido usar correlação fixando os valores de uma variável para estudar as alterações na outra.

B) Regressão simples

O que é regressão simples ?

Uma regressão simples examina a relação entre uma variável de resposta contínua (Y) e uma variável preditora (X).

Quando usar uma regressão simples ?

Use uma regressão simples quando possuir um Y contínuo e um único X. As condições a seguir deverão ser atendidas :

- X pode ser ordinal ou contínuo
- Na teoria, X deveria ser fixo. Na prática contudo, ele frequentemente pode variar
- Presume-se que qualquer variação aleatória na medição de X é desprezível se comparada à faixa na qual X é medido

Os valores-Y observados na amostra irão diferir daqueles previstos pelo modelo de regressão (a menos que todos os pontos caiam perfeitamente em uma reta). Essas diferenças são chamadas resíduos.

Antes de iniciar uma análise de regressão, deve-se verificar se as suposições sobre os resíduos são válidas :

- Eles são independentes (e, portanto, aleatórios)
- Eles estão distribuídos normalmente
- Eles tem uma variância constante ao longo de todos os valores de X

Por que usar uma regressão simples

Uma regressão simples pode ajudar a responder questões tais como:

- Quão bom é X para prever Y?
- Qual valor pode-se esperar para Y quando X é 20?
- Quanto Y irá mudar se aumentar X em uma unidade?

Por exemplo :

- Como a temperatura de processamento se relaciona com a resistência do aço?
- Quão resistente será o aço se ele for processado a uma temperatura em particular?
- Quão resistente será o aço se aumentarmos a temperatura em 100°C

Ajustando um modelo linear

Estamos com muitos problemas de deslocamento de tinta nas estruturas metálicas, fornecidas pela Keiper. O respectivo fornecedor, está realizando testes para determinar o efeito da taxa de agitação no nível de impurezas da tinta que é utilizada para pintar as peças.

```

187 #Fit Model
188 tinta<- read_excel("TINTA.xlsx")
189 head(tinta)
190 library("ggplot2")
191 ggplot(tinta, aes(x=`Taxa Agitação`, y=Impurezas)) + geom_point() + stat_smooth()
192 cor(tinta$`Taxa Agitação`, tinta$Impurezas)
193 cor.test(tinta$`Taxa Agitação`, tinta$Impurezas)
194 help ("lm")
195 model = lm(tinta$Impurezas ~ tinta$"Taxa Agitação")
196 ggplot(tinta, aes(x=`Taxa Agitação`, y=Impurezas)) + geom_point() + stat_smooth(method = lm)
197 summary(model)
198 confint(model)

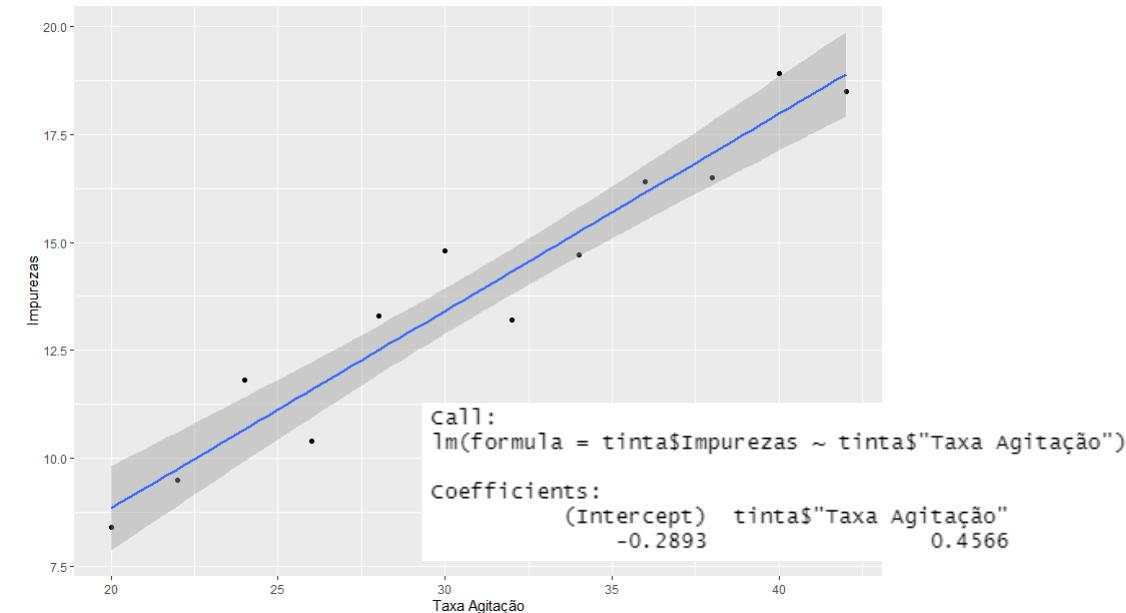
```

O que podemos ver no código acima:

1. Abrimos o arquivo tinta.xlsx
2. Motramos as primeiras linhas
3. Carregamos o ggplot2
4. Fazemos a curva de correlação
5. Calculamos a correlação e realizamos os testes
6. Geramos o modelo linear e salvamos em model
7. Plotamos a curva da regressão linear
8. Apresentamos o resultado do modelo
9. Apresentamos o intervalo de confiança

Agora vamos entender os resultados.

Interpretação dos resultados



Equação de Regressão

A equação de regressão relaciona a variável preditora (**Agitação**) a uma variável resposta (**Impurezas**): **Impurezas** : -0,289 +0,4566***Taxa Agitação**

A inclinção da linha de regressão, (0,4566), indica quanto uma mudança em **Impurezas** está associada com uma mudança em uma unidade de **Taxa Agitação**.

S

S é uma estimativa média ao redor da linha de regressão. S é a raiz quadrada de MSE. Quanto mais a equação for capaz de prever a resposta, menor será S.

Interpretação dos resultados

R² (R-Sq)

R² é a proporção da variabilidade da variável resposta que é explicada pela equação. Assim, 93,4% da variação em **Impurezas** pode ser explicada pela relação linear com **Taxa Agitação**.

Os valores aceitáveis para R² variam de acordo com o estudo. Por exemplo, engenheiros que estejam estudando relações químicas podem requerer um R² de 90% ou mais. Contudo, alguém que esteja estudando o comportamento humano (que é mais variável) pode ficar satisfeito com valores de R² muito mais baixos.

R² ajustado (R-Sq(adj))

R² ajustado não é sensível ao número de termos do modelo e é importante ao se comparar modelos com números diferentes de termos.

```
Call:
lm(formula = tinta$Impurezas ~ tinta$"Taxa Agitação")

Residuals:
    Min      1Q  Median      3Q     Max 
-1.1834 -0.5432 -0.3233  0.8333  1.3900 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.28928   1.22079  -0.237  0.817    
tinta$"Taxa Agitação" 0.45664   0.03844 11.880 3.21e-07 ***
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9193 on 10 degrees of freedom
Multiple R-squared:  0.9338, Adjusted R-squared:  0.9272 
F-statistic: 141.1 on 1 and 10 DF,  p-value: 3.211e-07
```

Interpretação dos resultados

Com base na análise ao lado (ANOVA), podemos retirar as seguintes informações :

Taxa Agitação (X) ; Impurezas (Y)

- Modelo de regressão : $Y = \beta_0 + \beta_1 X + \varepsilon$
- Modelo restrito : $Y = \beta_0 + \varepsilon$

O modelo restrito considera que as mudanças em Y se devem exclusivamente a erro aleatório (ε). Ele é equivalente a um modelo de regressão simples com uma inclinação (β_1) de zero. Assim, as hipóteses para a ANOVA são :

- H₀ : β_1 é igual a zero
- H_a : β_1 não é igual a zero

Interprete o valor-p (P) da seguinte maneira :

- Se o valor-p for menor que ou igual a α , rejeite H₀. O modelo de regressão explica significativamente mais variabilidade na resposta do que o modelo restrito β_1 não é igual a zero.
- Se o valor-p for maior que α , não poderá rejeitar H₀. β_1 não é significativamente diferente de zero.

Conclusão :

Usando um α de 0,05, pode-se rejeitar o modelo restrito, mais simples, e declarar que **Taxa Agitação** tem um efeito linear significativo sobre **Impurezas**.

Adicionando faixas de confiança e previsão

Faixas de confiança e previsão

Desejamos nos sentir confiantes de que os pontos médios e individuais da variável Y, **Impurezas**, caem dentro de certos limites de variabilidade.

Resíduos e ajustados

Resíduos são as diferenças entre os valores ajustados do modelo e os valores observados. Valores ajustados são as estimativas da resposta média para cada nível da variável independente.

Usaremos o nível de confiança de 95%

	2.5 %	97.5 %
(Intercept)	-3.009365	2.4308107
tinta\$"Taxa Agitação"	0.370997	0.5422898

Interpretação dos resultados

Intervalos de Confiança :

O intervalo de confiança de 95% define uma faixa provável de valores para a média populacional de Y. Para qualquer valor dado de X, pode-se ter 95% de confiança de que a média populacional para Y está entre as linhas indicadas.

Intervalos de Previsão :

O intervalo de previsão de 95% define uma faixa provável de valores de Y para observações individuais. Para qualquer valor dado de X, pode-se ter 95% de confiança de que o valor correspondente de Y para uma observação estará entre as linhas indicadas.

Considerações finais

Conclusões práticas

Uma simples análise de regressão linear revelou que um aumento nas taxas de agitação está associado a crescentes níveis de impurezas na tinta.

A inclinação da equação de regressão indica que, quando a taxa de agitação é aumentada em 1 rpm, o nível de impurezas aumenta em 0,4566.

Pode-se usar a equação para determinar qual será o nível de impureza para diferentes taxas de agitação. Contudo, a equação só é válida para a faixa de dados em que as amostras foram coletadas (taxas de agitação entre 20 e 42).

Considerações estatísticas

Não pode-se usar análise de regressão para avaliar se mudanças nas preditoras causam mudanças na resposta, a não ser no caso em que os valores das preditoras são fixados em níveis pré-determinados em um experimento controlado. Se os valores das preditoras puderem variar aleatoriamente, outros fatores poderão influenciar tanto as preditoras quanto a resposta.

Não aplique resultados de regressão aos valores de X que estejam fora da faixa de amostras. Por exemplo, não use a equação de regressão derivada deste exemplo para prever níveis de impureza para uma taxa de agitação de 100, pois a taxa de agitação mais elevada envolvida na análise é 42. A relação entre **Taxa Agitação** e **Impurezas** pode ser muito diferente para taxas de agitação acima de 42.

Fique alerta para **outliers** quando estiver adotando procedimentos de regressão. Alguns **outliers** (chamados pontos de alta influência) Tem um grande efeito sobre o cálculo da reta de regressão por mínimos quadrados. Em tais casos, a reta pode não mais representar bem o resto dos dados.



11

Análise de Variância

Nível: Avançado

One-way ANOVA

O que é one-way ANOVA ?

O procedimento one-way ANOVA (análise de variância) é uma generalização do teste t de amostras independentes. Ao contrário do teste t, contudo, one-way ANOVA pode ser usado para analisar as médias de mais de 2 grupos (amostras) ao mesmo tempo.

A idéia básica da ANOVA é a seguinte :

- A variação dentro de cada grupo é devida apenas ao erro aleatório;
- Portanto, se a variação entre grupos for similar à variação em cada grupo, é provável que as médias entre os grupos sejam diferentes apenas devido ao erro aleatório;
- Contudo, se a variação entre grupos for grande em relação à variação dentro do grupo, é provável que as diferenças entre as médias dos grupos sejam causadas por diferenças nos níveis dos fatores.

Quando usar one-way ANOVA ?

Use one-way ANOVA (também chamada de ANOVA de fator único) quando tiver respostas contínuas para 2 ou mais níveis fixos de um único fator.

Antes de aceitar os resultados de uma ANOVA, verifique se as seguintes suposições sobre os resíduos são válidas para os dados :

- Eles devem ser independentes (e, portanto, aleatórios);
- Eles não devem se desviar de forma substancial de uma distribuição normal;
- Eles devem ter variância constante ao longo de todos os níveis de fatores.

Por que usar one-way ANOVA ?

One-way ANOVA pode ajudar a responder questões tais como :

Existem diferenças entre os produtos dos fornecedores ?

Existem diferenças entre grupos de tratamentos ?

Por exemplo :

As resistências das amostras de plástico de 4 fornecedores são diferentes?

A combustão se torna mais eficiente se utilizar o aditivo de combustível A, o aditivo B ou nenhum aditivo ?

Executando o one-way ANOVA

Use um one-way ANOVA para comparar os tempos médios de montagem dos diferentes tipos de carpetes utilizados em nossos bancos automotivos e crie gráficos para visualizar os dados.

One-way

```
207 #Anova
208 carpete<- read_excel("CARPETO.xlsx")
209 fit<- aov(carpete$Tempo_Seg ~ carpete$Carpete_Tipo)
210 layout(matrix(c(1,2,3,4),2,2))
211 plot(fit)
212 summary(fit)
```

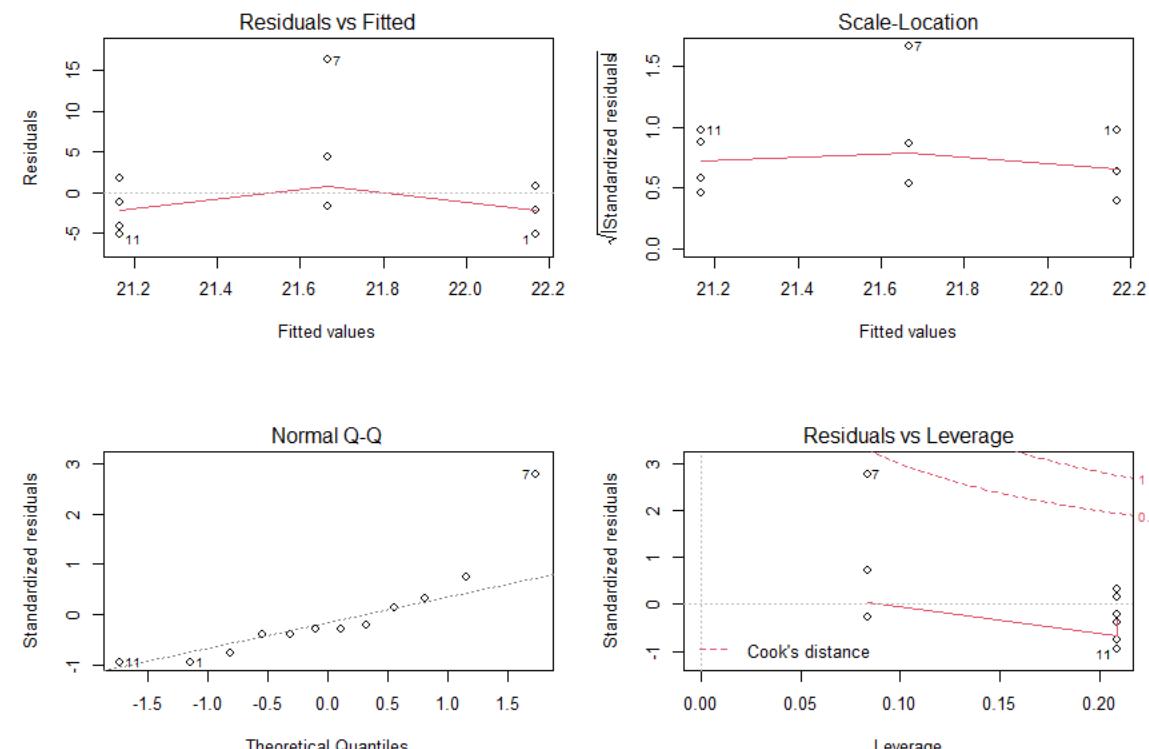
Acima, a função aov cria o modelo da ANOVA e salva na variável fit.

Depois mudamos o layout da área de plotagem para 4 quadrantes, 2 por 2.

Plotamos o modelo e, apresentamos os resultados.

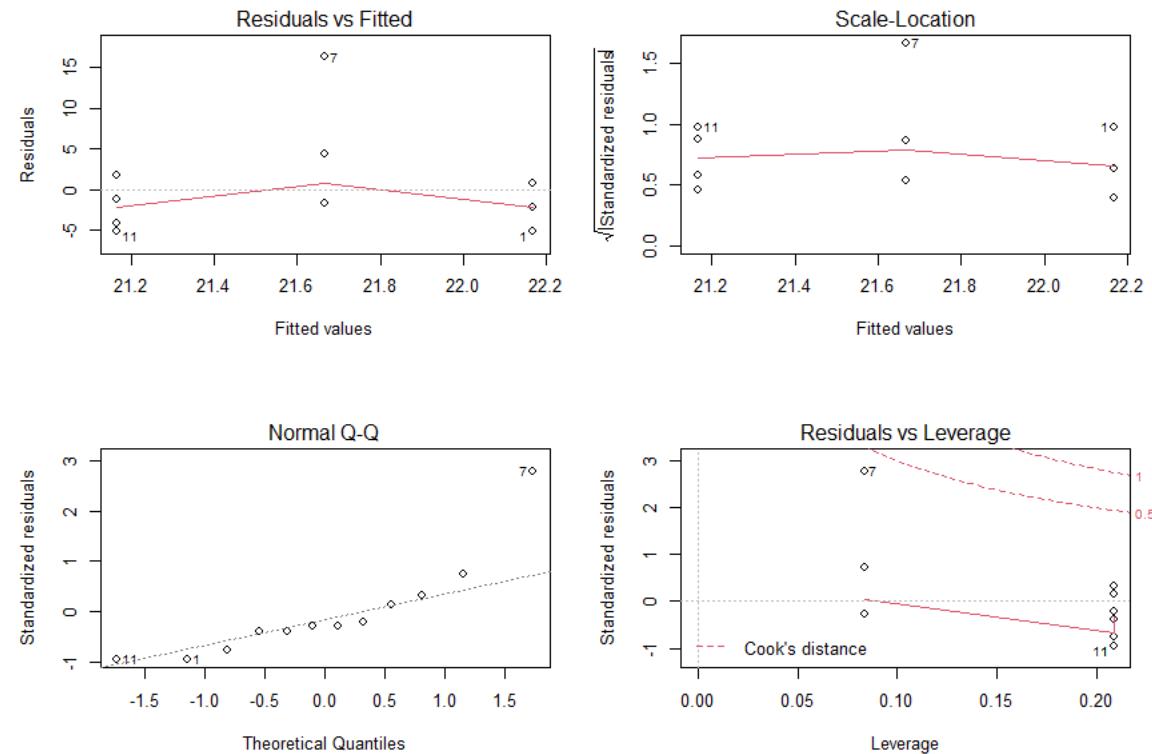
	df	Sum Sq	Mean Sq	F value	Pr(>F)
carpete\$Carpete_Tipo	1	2.0	2.00	0.053	0.822
Residuals	10	376.7	37.67		

Interpretação dos resultados



O individual value plot revela que o grupo 2 contém uma única observação com um valor incomumente alto. Com apenas 4 observações em cada grupo, um ponto outlier dessa natureza pode ter um grande efeito na média e no desvio padrão da amostra.

Interpretação dos resultados



Outliers desse tipo podem ser o resultado de variação aleatória ou podem indicar que alguma coisa estranha está ocorrendo em seu processo. Investigue outliers para determinar as causas sempre que possível.

Para análise presente, pressuponha que todas as observações são válidas.

Interpretação dos resultados

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
carpete\$Carpete_Tipo	1	2.0	2.00	0.053	0.822
Residuals	10	376.7	37.67		

Análise da Variância

A 1ª fila na tabela Analysis of Variance (Análise de Variância) contém as estatísticas associadas com o fator, **Carpete_Tipo**. A fila seguinte contém as estatísticas associadas com o erro aleatório, **Error**.

Graus de liberdade

Os graus de liberdade (**DF**) relacionam-se ao número de valores usados para calcular a soma dos quadrados (**SS**) para cada fonte de variação.

Soma dos quadrados

A soma dos quadrados (**SS**) mensura a quantidade de variabilidade que cada fonte de variação contribui para os dados. Observe que a quantidade total de variabilidade dos dados (**SS Total**, 376,7) é igual a **SS** para **Carpete_Tipo** (114,7) mais a **SS** para **Error** (264,0).

Interpretação dos resultados

	DF	Sum Sq	Mean Sq	F value	Pr(>F)
carpete\$Carpete_Tipo	1	2.0	2.00	0.053	0.822
Residuals	10	376.7	37.67		

Quadrado Médio

O quadrado médio (**MS**) para cada fonte é igual a **SS** dividida por **DF**.

- A **MS** para o fator é uma estimativa da variabilidade média entre os grupos.
- A **MS** para erro é uma estimativa da variabilidade média dentro do grupo.

Estatística F

F é a razão entre a variabilidade atingida ao fator a variabilidade atribuída ao erro. Ela é calculada como a **MS** para o fator (**Carpete_Tipo**) dividida pela **MS** para o erro.

- Quando as diferenças entre as médias dos níveis do fator são similares às diferenças entre observações dentro de cada nível, **F** estará próximo de 1.
- Se a variabilidade entre as médias dos níveis dos fatores for maior do que a variabilidade entre observações dentro do nível do fator, **F** será maior que 1.

Interpretação dos resultados

Valor-p

O valor-p é a probabilidade de **F** ser tão grande quanto é, mesmo com seu fator não tendo qualquer efeito. Ter um **F** grande sugere que as médias dos níveis dos fatores são mais diferentes do que seria se fosse devido ao acaso; portanto, o **valor-p** é pequeno.

Use o **valor-p** para testar as seguintes hipóteses :

- Ho (hipótese nula) : todas as médias dos níveis dos fatores são as mesmas
- Ha (hipótese alternativa) : as médias dos níveis dos fatores são diferentes.

Conclusão

Uma vez que o **valor-p** é maior que a (0,05), não podemos rejeitar Ho. Não há evidências suficientes para sugerir que as médias dos níveis são diferentes.

Considerações finais

Conclusões práticas

A análise dos dados do carpete não revelou quaisquer diferenças no tempo de aquecimento entre os lotes. Contudo, houve problemas potenciais no estudo:

- Outlier – um valor pareceu ser um outlier e deve ser investigado;
- Padrões – 2 carpetes consecutivos apresentaram tempos de montagem menores que o normal. Também pode haver evidência de um aumento sistemático nos tempos de montagem dos primeiros 4 carpetes testados;
- Baixo poder – com base em um σ de 5,416, o poder do teste para detectar uma diferença de 7 segundos (ao nível α de 0,05) é de apenas 0,2642. Isso é menos que 27% de chance de detectar em uma diferença. De fato, com um poder de 0,80, deveríamos ter uma diferença de mais de 14 segundos.

Com base nesses resultados, talvez o melhor caminho a se tomar seja garantir, primeiro, que o processo esteja sob controle e, depois, coletar amostras maiores e executar o teste novamente. Se σ for reduzido em 3,0 e coletarmos amostras de 6 carpetes de cada tipo, será possível detectar uma diferença de 7 segundos com um poder de 0,9133.

Considerações estatísticas

É preferível comparar fatores de múltiplos níveis com uma única ANOVA do que comparar 2 níveis de uma vez com 2 testes t separados. Isso porque executar testes extras aumenta as chances de um erro tipo I (rejeitar H_0 quando H_0 é, de fato, verdadeira).

A suposição de independência é crítica. Se as observações forem afetadas sistematicamente por fatores diversos daquele que está estudando, os resultados da one-way ANOVA não terão significado.

Geralmente, a suposição de normalidade não é crucial, especialmente se as amostras forem grandes.



12

Experimentos Fatoriais Completos

Nível: Mestre

Experimentos Fatoriais

O que são experimentos fatoriais ?

Os experimentos fatoriais permitem o estudo simultâneo dos efeitos de vários fatores em um processo. Variar os níveis dos fatores simultaneamente, ao invés de um de cada vez, é eficiente em termos de tempo e custo e permite o estudo das interações entre os fatores.

Quando usar experimentos fatoriais ?

Os experimentos fatoriais são utilizados para :

- estimar de forma eficiente o efeito de cada fator sobre a resposta;
- estimar os efeitos de interações entre dois ou mais fatores na resposta
- testar a curvatura da resposta pela inclusão de pontos centrais no experimento

Por que usar experimentos fatoriais ?

Os experimentos fatoriais são utilizados para responder a questões tais como :

- quais variáveis tem maior influência na resposta ?
- que configurações de fatores irão otimizar a resposta ?

Por exemplo :

- qual o impacto do corte, da resistência do metal e do ângulo de corte na vida útil de uma ferramenta para cortar metais ?
- que configurações de adoçante, taxa de xarope, nível de carbonação e temperatura maximizam a pontuação de sabor de um novo refrigerante ?

Experimentos Fatoriais Completos

Experimentos fatoriais completos

O que são experimentos fatoriais completos ?

Em um experimento fatorial completo, as respostas são medidas em todas as combinações dos níveis dos fatores experimentais. As combinações dos níveis dos fatores representam as condições nas quais as respostas são mensuradas. Cada combinação experimental é um ensaio e cada resposta medida é uma observação. O conjunto de todos os ensaios é o experimento.

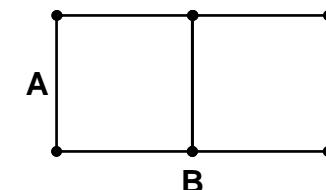
O diagrama a seguir mostra experimentos de dois e três fatores. Os pontos representam uma combinação única de níveis de fatores. Por exemplo, no experimento de dois fatores, o ponto no canto inferior esquerdo representa o ensaio experimental quando o Fator A é ajustado ao seu nível inferior e o Fator B também é ajustado ao seu nível inferior.

Quando usar experimentos fatoriais completos ?

O MINITAB oferece dois tipos de experimentos fatoriais completos :

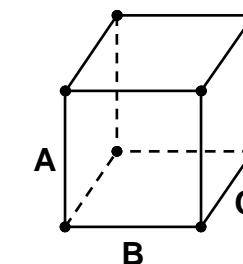
- Experimento fatorial de dois níveis (experimento 2^k), que deve ser utilizado quando cada fator experimental tiver apenas dois níveis
- Experimento fatorial completo geral (General Linear Model), que deve ser utilizado quando qualquer fator experimental tiver mais de dois níveis. Por exemplo, o Fator A pode ter dois níveis, o Fator B pode ter três níveis e o Fator C pode ter cinco níveis.

Dois fatores



Dois níveis do Fator A
Três níveis do Fator B

Três fatores



Dois níveis de cada fator

Por que usar experimentos fatoriais completos ?

Os experimentos fatoriais completos são utilizados para responder a questões tais como :

- quais variáveis têm a maior influência na resposta ?
- existem interações entre dois ou mais fatores que influenciam a resposta ?
- quais configurações de fatores irão otimizar a resposta ?

Por exemplo :

- qual o impacto do tipo de vidro e do tipo de fósforo no brilho de um tubo de televisão
- como a interação entre a temperatura e o açúcar afeta a maciez de um chocolate
- que forma de bocal, para uma ferramenta de corte com jato d'água, irá minimizar o tempo necessário para cortar uma chapa padrão de metal ?

Criando um experimento fatorial

Experimento

Com 2 fatores, a única escolha é um experimento fatorial completo, que contém 4 ensaios (1 para cada combinação de níveis). Fazendo-se 4 réplicas deste experimento, resulta em 16 ensaios.

Níveis dos fatores

Por default, o MINITAB nomeia os fatores em ordem alfabética e os configura com códigos : low (baixo) = -1, high (alto) = +1.

Utilizando-se valores não codificados para os níveis alto e baixo do fator, podem-se obter os coeficientes do modelo em unidades não codificadas e gráficos rotulados de maneira apropriada.

Base para a geração de dados aleatórios

Por default, o RStudio aleatoriza a ordem dos ensaios. A aleatorização mistura a ordem dos ensaios de tal forma que a variação sistemática não é confundida com os efeitos dos fatores.

Criar Experimentos Fatoriais

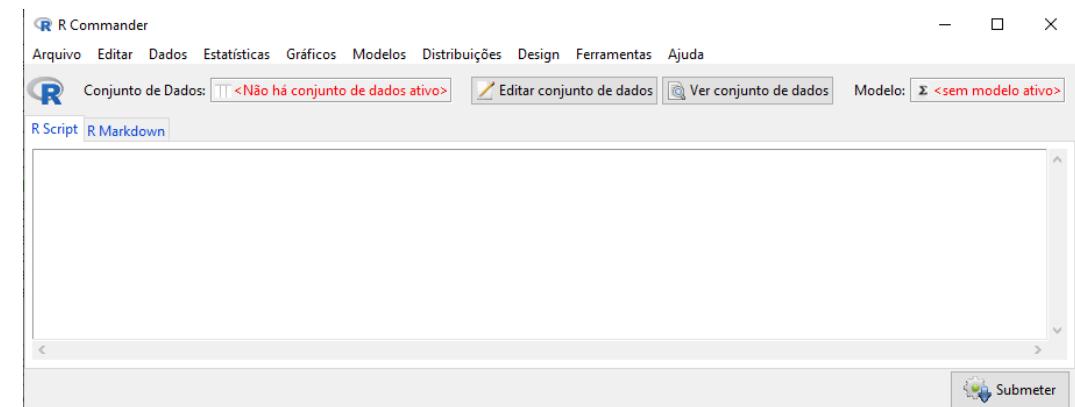
```
216 #DoE  
217 adesao<- read_excel("ADESAO.xlsx")  
218 view(adesao)  
219 library("RcmdrPlugin.DOE")  
220 resposta=adesao$TempCol  
...
```

Caso você queira simplesmente analisar um DoE já realizado, os passos acima podem ser feitos assim, na sequência. Mas caso você ainda vá fazer o DoE, antes você precisará gerar a lista, fazer o ensaio e depois começar carregando o arquivo. Eu vou mostrar...

Ao digitar os comandos acima, você já carregou o arquivo com o ensaio e as respostas, já abriu o commander (para o DoE o Rstudio abre um Plugin ☺)

E por fim já colocou as respostas em uma variável de ambiente!

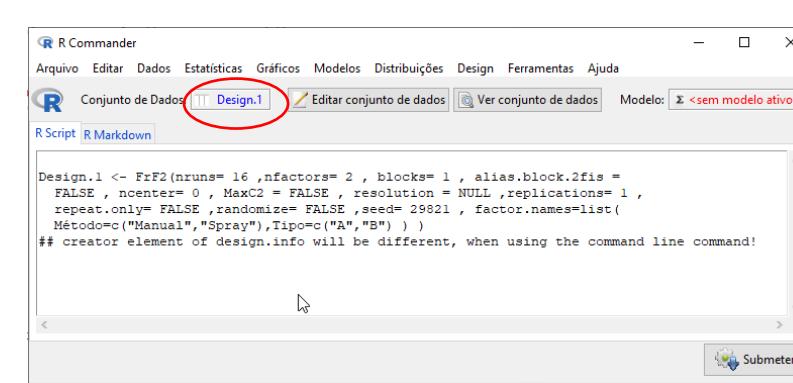
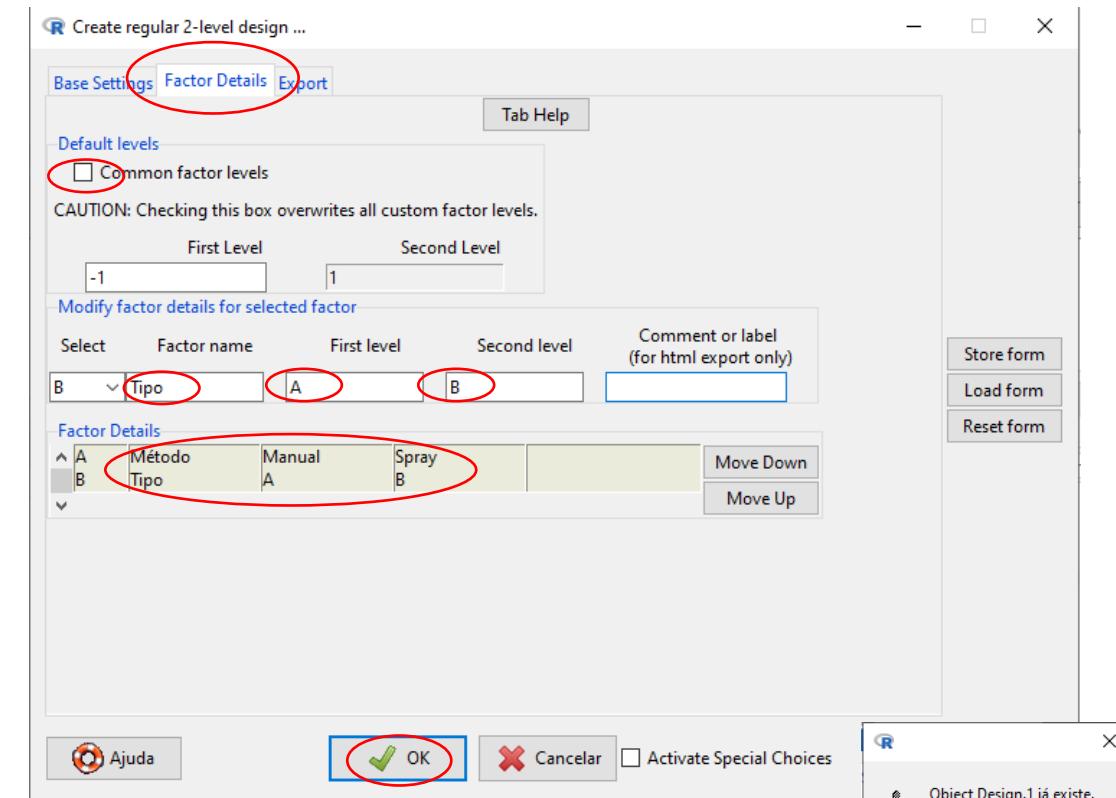
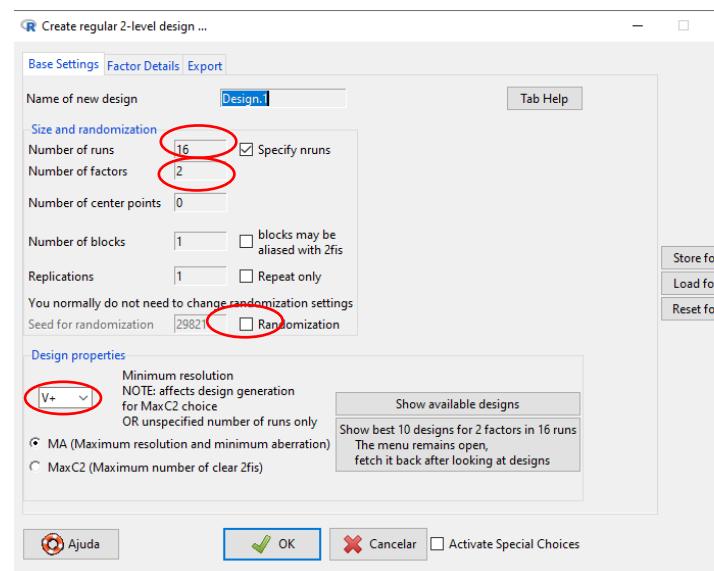
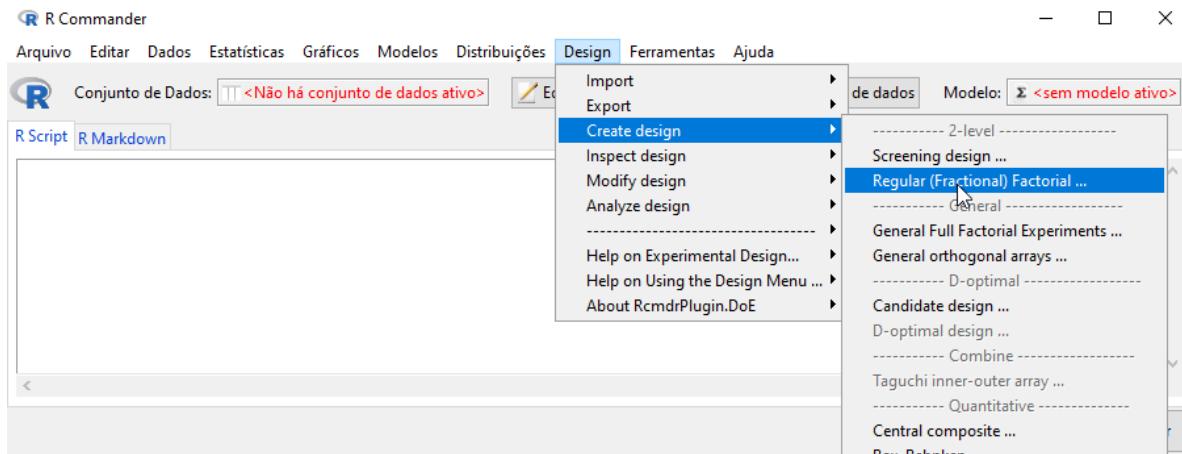
Vamos nessa, para começar essa tela irá se abrir:



Experimentos Fatoriais Completos

Criando um experimento fatorial

Vamos ao commander:

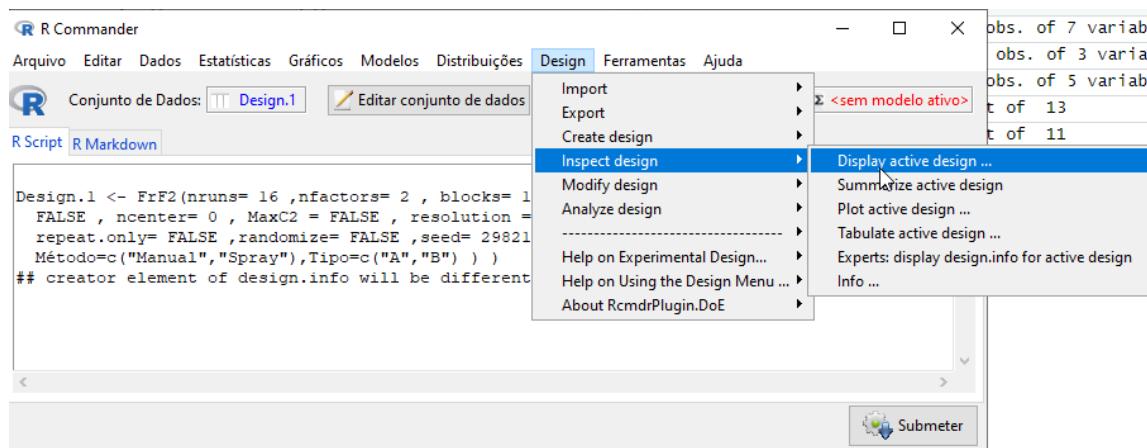


Experimentos Fatoriais Completos

Inspecionando o modelo

Vamos ao commander:

O modelo já foi criado. Mas como ele está? E se agora eu quiser imprimir a folha de ensaios para realizar os ensaios?



```
Rcmdr> print( Design.1 )
      run.no run.no.std.rp Método Tipo Blocks
1       1           1.1 Manual   A    .1
2       2           2.1 Spray    A    .1
3       3           3.1 Manual   B    .1
4       4           4.1 Spray    B    .1
5       5           1.2 Manual   A    .2
6       6           2.2 Spray    A    .2
7       7           3.2 Manual   B    .2
8       8           4.2 Spray    B    .2
9       9           1.3 Manual   A    .3
10     10          2.3 Spray    A    .3
11     11          3.3 Manual   B    .3
12     12          4.3 Spray    B    .3
13     13          1.4 Manual   A    .4
14     14          2.4 Spray    A    .4
15     15          3.4 Manual   B    .4
16     16          4.4 Spray    B    .4

class=design, type= full factorial
NOTE: columns run.no and run.no.std.rp are annotation,
not part of the data frame
```

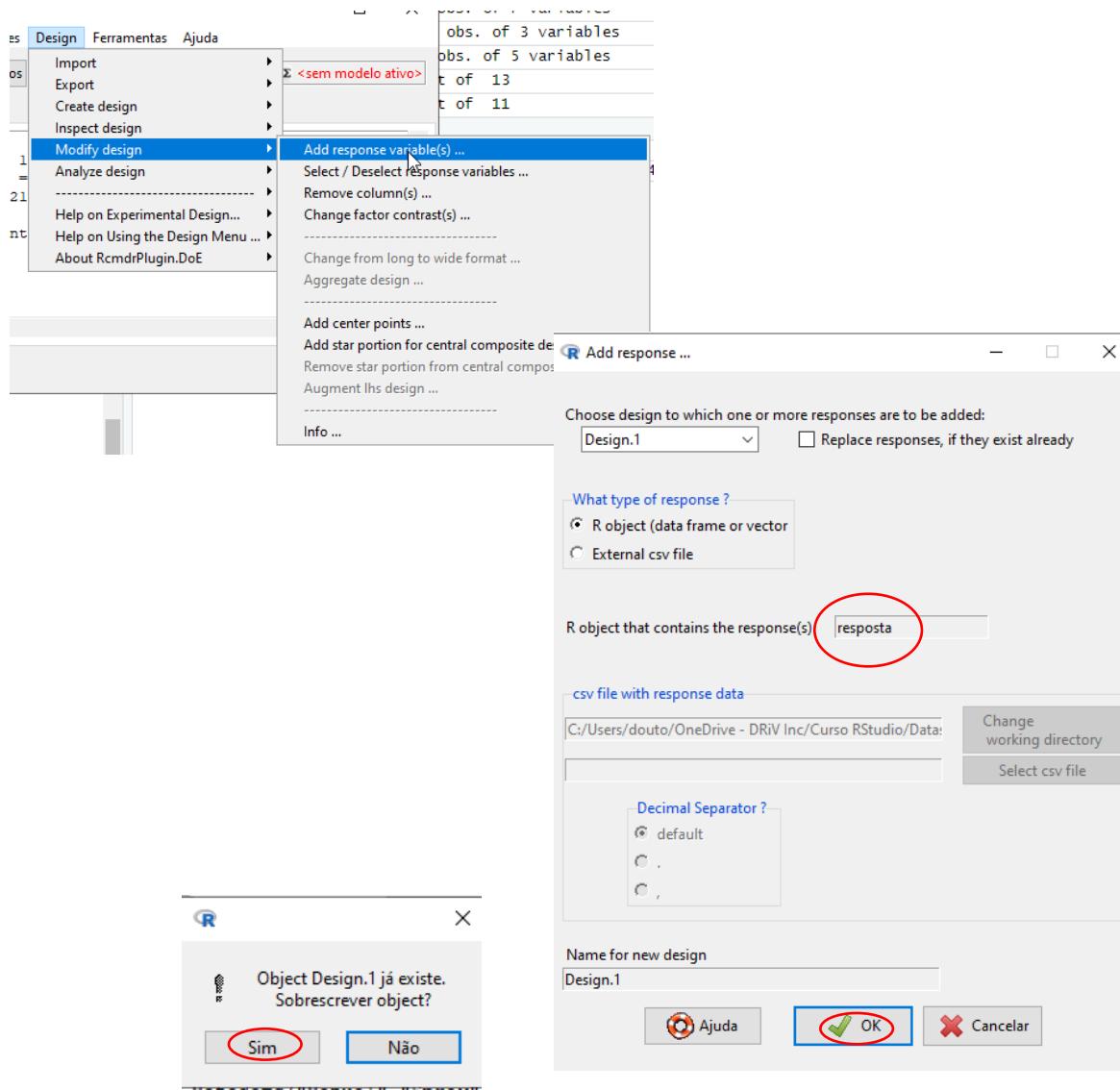
A resposta sai lá
No console do Rstudio:

Visualizando dados

Antes, uma outra forma de ver se está tudo bem com os dados é editar ou visualizar em forma de tabela

Experimentos Fatoriais Completos

Carregando as respostas



Visualizando dados

Antes, uma outra forma de ver se está tudo bem com os dados é editar ou visualizar em forma de tabela

The screenshot shows the R Commander Editor de dados window. The top bar shows 'Arquivo', 'Editar', 'Dados', 'Estatísticas', 'Gráficos', 'Modelos', 'Distribuições', 'Design', 'Ferramentas', and 'Ajuda'. The 'Modelos' tab is selected. The main area shows a data table with columns: rowname, Método, Tipo, Bloco, and resposta. The 'resposta' column is circled in red. The table contains 16 rows of data. At the bottom are buttons for 'Ajuda', 'OK' (highlighted with a red circle), and 'Cancelar'.

rowname	Método	Tipo	Bloco	resposta
1	1	Manual	A	4.52
2	2	Spray	A	4.37
3	3	Manual	B	5.14
4	4	Spray	B	4.88
5	5	Manual	A	4.29
6	6	Spray	A	4.25
7	7	Manual	B	5.05
8	8	Spray	B	4.75
9	9	Manual	A	4.48
10	10	Spray	A	4.43
11	11	Manual	B	4.95
12	12	Spray	B	4.71
13	13	Manual	A	4.60
14	14	Spray	A	4.55
15	15	Manual	B	4.89
16	16	Spray	B	4.91

Experimentos Fatoriais Completos

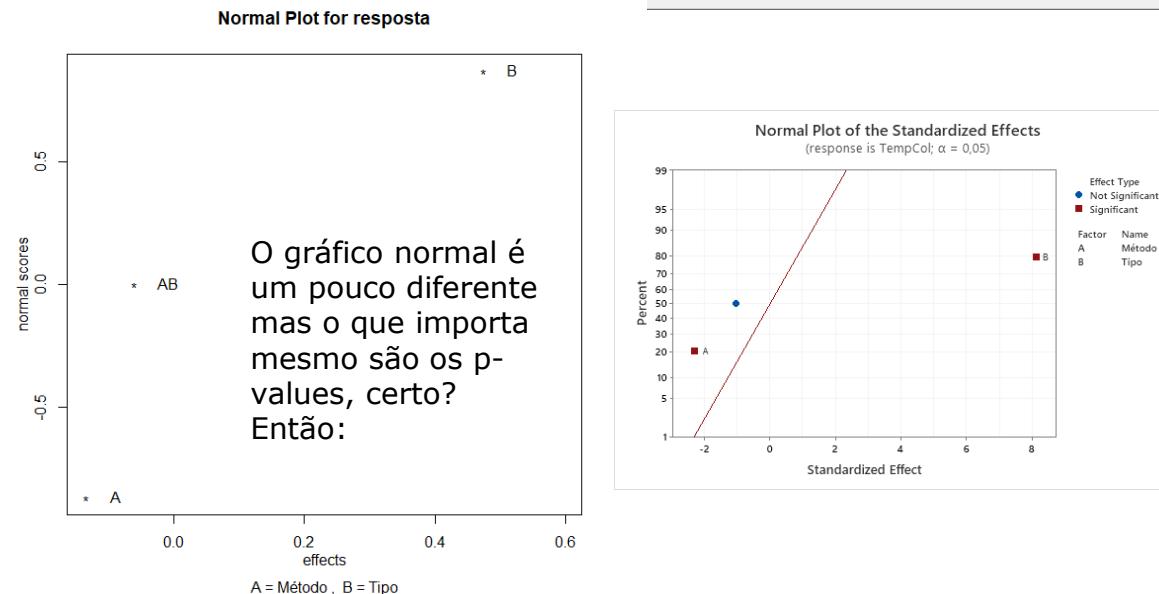
Carregando as respostas

Rcmdr menu: Design > Analyze design > Effects (Half) Normal Plots ...

Effects plots for 2-level factors dialog:

- Response to be analysed (select one): resposta
- Half normal plot?
- Label effects with codes instead of names?
- Enter significance level for labelling: 0.1
- Label significant effects only?

Buttons: Ajuda (Help), OK (highlighted), Cancelar (Cancel).



Gerando o modelo linear

Rcmdr menu: Design > Default linear model ...

Choose Response and Degree dialog:

- Response to be analysed (select one): resposta
- degree (positive integer): NULL

Buttons: Ajuda (Help), OK (highlighted), Cancelar (Cancel).

Linear model for experimental designs dialog:

- Variáveis (clique-duplo para fórmula):
 - Blocks [fator]
 - Método [fator]
 - resposta
 - Tipo [fator]
- Fórmula do modelo:
 - Operadores (clique na fórmula): +, *, :, /, %in%, -, ^, (,)
 - Splines/Polinômios: (selecionar variável e clicar)
 - B-spline
 - natura spline
 - polinômio ortogonal
 - polinômio bruto
 - GL para splines
 - graus para polinômios

Buttons: Ajuda para fórmula do modelo (Help for formula), OK (highlighted), Cancelar (Cancel).

Modelo LinearModel.1 já existe. Sobre escrever modelo? dialog:

Sim (Yes), Não (No).

Carregando as respostas

R Commander interface showing the 'Design.1' dataset selected in the 'Conjunto de Dados' dropdown and 'LinearModel.' selected in the 'Modelo:' dropdown. The R Script pane contains the following code:

```
DanielPlot(Design.1, code=TRUE, autolab=TRUE, alpha=0.1, half=FALSE,
           response="resposta")
DanielPlot(Design.1, code=TRUE, autolab=TRUE, alpha=0.1, half=TRUE,
           response="resposta")
DanielPlot(Design.1, code=TRUE, autolab=TRUE, alpha=0.05, half=TRUE,
           response="resposta")
editDataset(Design.1)
LinearModel.1 <- lm(resposta ~ (Método + Tipo)^2, data=Design.1)
summary(LinearModel.1)
```

The 'Submeter' button is visible at the bottom right.

Veja que agora o modelo está completo e com a resposta:

```
Call:
lm.default(formula = resposta ~ (Método + Tipo)^2, data = Design.1)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.18250 -0.07250  0.01875  0.07500  0.15000 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 4.67312   0.02919 160.079 < 2e-16 ***
Método1    -0.06688   0.02919  -2.291   0.0409 *  
Tipo1       0.23687   0.02919   8.114  0.00000325 ***
Método1:Tipo1 0.03062   0.02919  -1.049   0.3148    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1168 on 12 degrees of freedom
Multiple R-squared:  0.8575, Adjusted R-squared:  0.8218 
F-statistic: 24.06 on 3 and 12 DF,  p-value: 0.00002304
```

Comparando com o minitab

Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		4,6731	0,0292	160,08	0,000	
Método		-0,1337	-0,0669	0,0292	-2,29	0,041
Tipo		0,4737	0,2369	0,0292	8,11	0,000
Método*Tipo		-0,0612	-0,0306	0,0292	-1,05	0,315

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,116771	85,75%	82,18%	74,66%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	3	0,98432	0,32811	24,06	0,000
Linear	2	0,96931	0,48466	35,54	0,000
Método	1	0,07156	0,07156	5,25	0,041
Tipo	1	0,89776	0,89776	65,84	0,000
2-Way Interactions	1	0,01501	0,01501	1,10	0,315
Método*Tipo	1	0,01501	0,01501	1,10	0,315
Error	12	0,16362	0,01364		
Total	15	1,14794			

Regression Equation in Uncoded Units

$$\text{TempCol} = 4,6731 - 0,0669 \text{ Método} + 0,2369 \text{ Tipo} - 0,0306 \text{ Método*Tipo}$$

Ou seja, tudo igual!

Interpretando os resultados

Os efeitos anteriormente calculados utilizando a tabela de médias encontram-se listados na tabela de efeitos estimados e coeficientes.

Os efeitos são :

- **4,67312** para a constante
- **-0,06688** para método de aplicação
- **0,23687** para tipo de adesivo
- **-0,03062** para a interação método de aplicação por tipo de adesivo

Pode-se determinar quais grupos de termos são estatisticamente significativos ao nível $\alpha = 0,05$, usando a tabela de análise de variância :

- Há pelo menos um efeito principal significativo ($p = 0,000$);
- O efeito da interação não é significativo ($p = 0,315$).

Com base nas informações do experimento, devemos eliminar os itens não significativos (neste caso, a interação AB) e reajustar o modelo.

```

call:
lm.default(formula = resposta ~ (Método + Tipo)^2, data = Design.1)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.18250 -0.07250  0.01875  0.07500  0.15000 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.67312   0.02919 160.079 < 2e-16 ***
Método1     -0.06688   0.02919  -2.291   0.0409 *  
Tipo1        0.23687   0.02919   8.114  0.00000325 ***
Método1:Tipo1 -0.03062   0.02919  -1.049   0.3148  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1168 on 12 degrees of freedom
Multiple R-squared:  0.8575,    Adjusted R-squared:  0.8218 
F-statistic: 24.06 on 3 and 12 DF,  p-value: 0.00002304

```

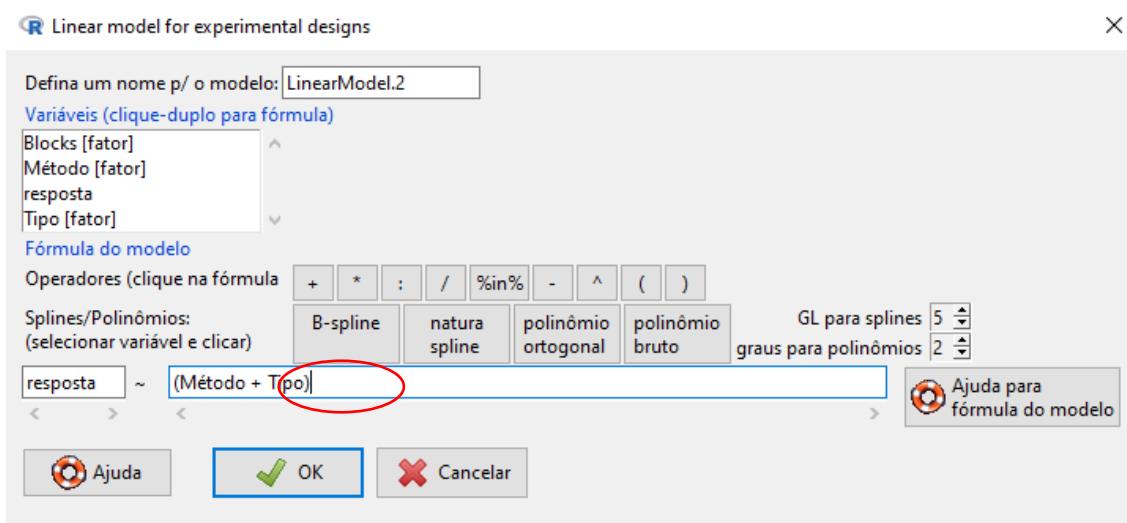
Reduzindo o modelo

Removendo termos do modelo

Para remover o termo de interação desse modelo, deve-se adotar um dos procedimentos a seguir :

- em **Include terms in the model up through order** (Incluir termos no modelo acima através da ordem), escolher **1**.
- destacar **AB** na caixa **Selected Terms** (Termos Selecionados) e clicar em
- dar um toque duplo no termo **AB** na caixa **Selected Terms**.

Os termos restantes (**A** e **B**) serão utilizados para explicar a variabilidade no tempo de adesão do hot melt.



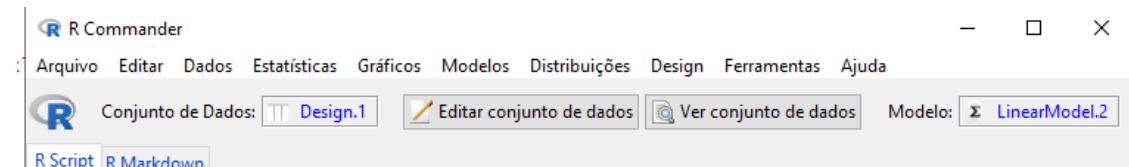
Interpretando os resultados

Veja que tiramos o $\wedge 2$ da fórmula, logo a interação não mais aparece no resultado:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.67312   0.02931 159.463 < 2e-16 ***
Método1     -0.06688   0.02931  -2.282    0.04 *  
Tipo1        0.23687   0.02931   8.083 0.000002 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

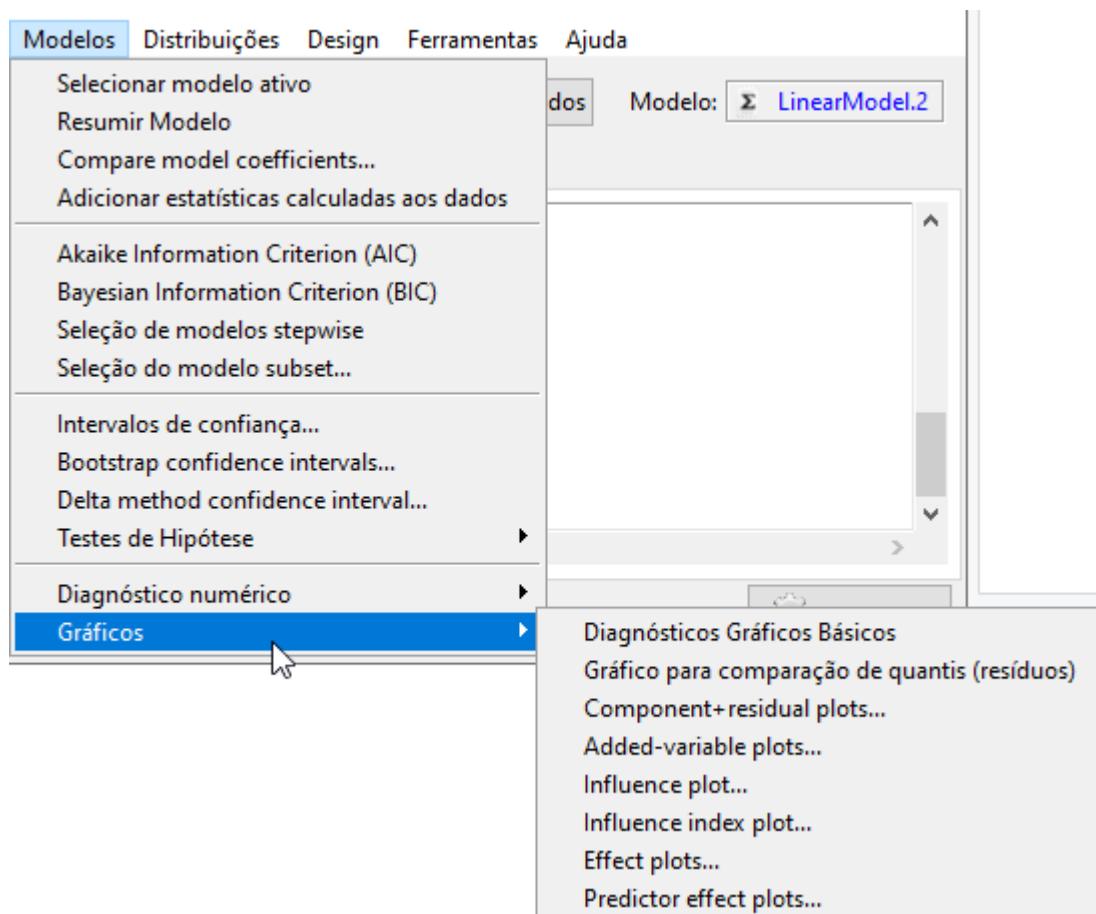
Residual standard error: 0.1172 on 13 degrees of freedom
Multiple R-squared:  0.8444,    Adjusted R-squared:  0.8205 
F-statistic: 35.27 on 2 and 13 DF,  p-value: 0.000005601
```

Outras análises podem ser feitas graficamente, vamos explorar os menus do Commander para mostrar mais.

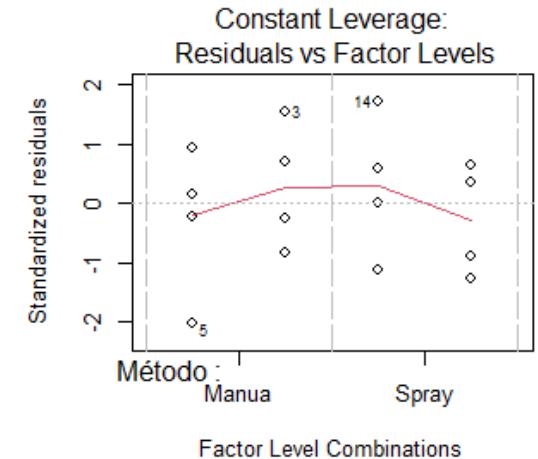
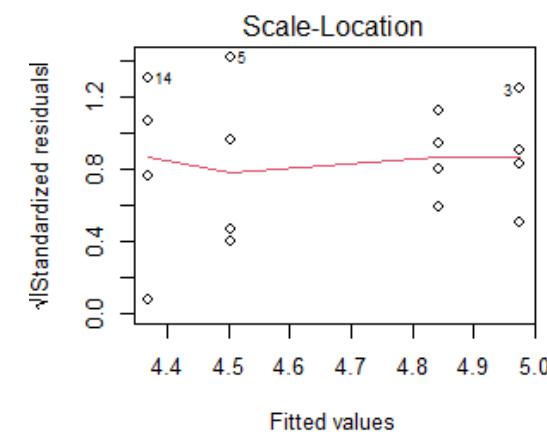
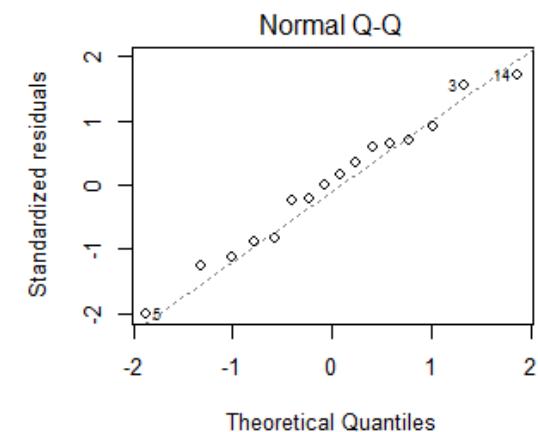
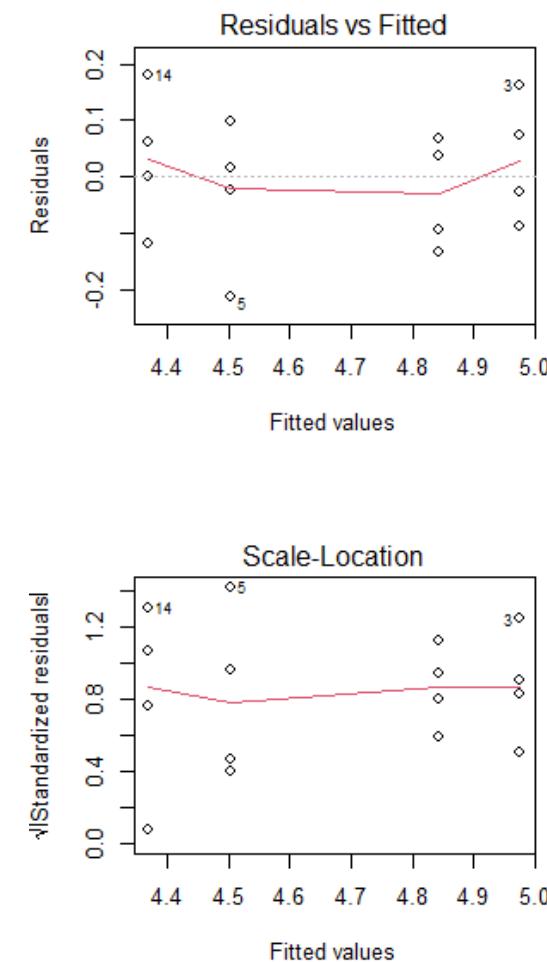


Analisando o modelo de forma gráfica

Para começar vamos analisar o menu gráficos da aba Modelos:



lm.default(resposta ~ (Método + Tipo))



Interpretando os resultados

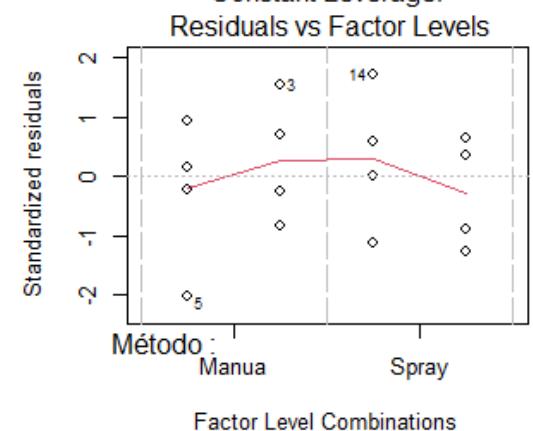
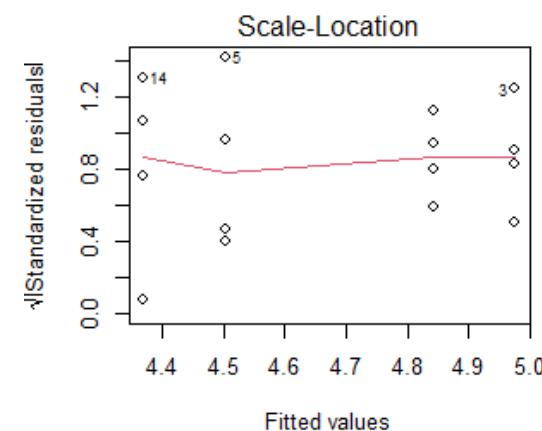
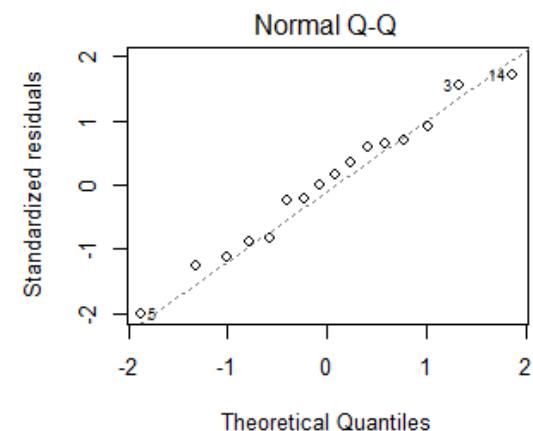
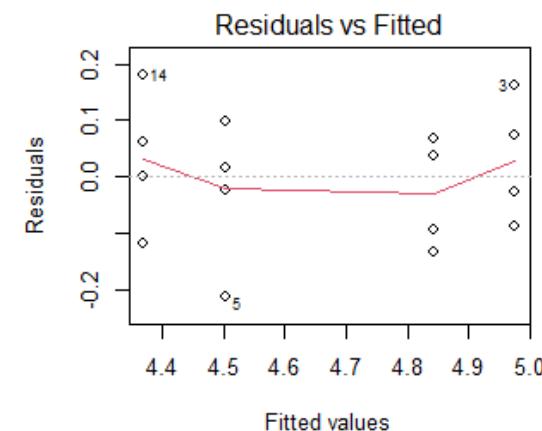
Utiliza-se o gráfico dos resíduos versus os valores ajustados para verificar se os seguintes pressupostos são satisfeitos :

- variância constante ao longo de todos os níveis de fatores
- não existe nenhum ponto discrepante dos dados

A tabela abaixo resume os padrões típicos possíveis de serem encontrados. A ocorrência desses padrões em um gráfico indica que um dos pressupostos foi violado.

Este padrão ...	Indica ...
dispersão em funil ou desigual dos resíduos ao longo de diferentes valores ajustados	a variância de seus resíduos não é constante
um ponto que se encontra muito longe de zero	um ponto discrepante
Para os dados de tempo de colagem, o gráfico dos resíduos versus os valores ajustados mostra uma dispersão aleatória, indicativa de uma variância constante dos resíduos.	

lm.default(resposta ~ (Método + Tipo))



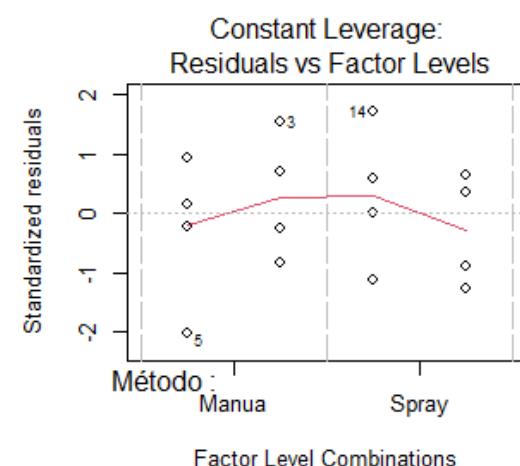
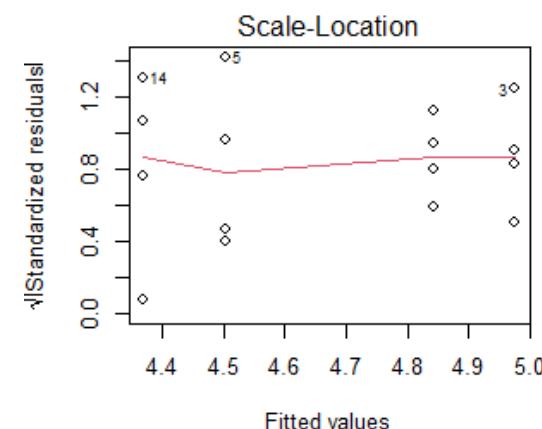
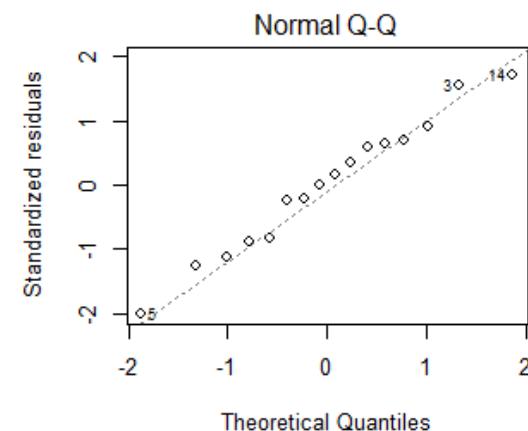
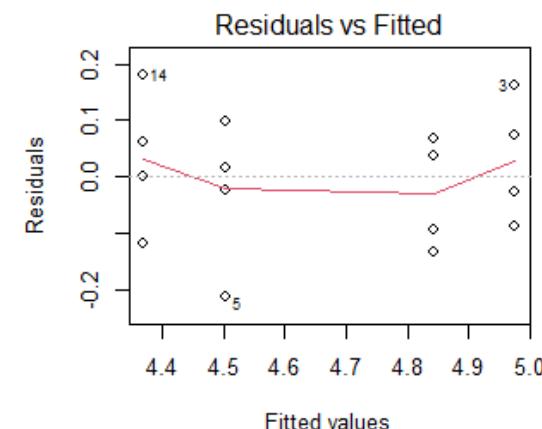
Interpretando os resultados

Utiliza-se o gráfico dos resíduos versus a ordem dos dados para verificar se os residuais são independentes.

- se houver um efeito devido à ordem de coleta dos dados, os resíduos não estarão dispersos aleatoriamente ao redor de zero. Neste caso, será possível detectar um padrão no gráfico;
- se não houver um efeito devido à ordem de coleta dos dados, os resíduos estarão dispersos aleatoriamente em torno de zero.

Para os dados de tempo de colagem, não há evidência de um efeito de tempo ou ordem.

`lm.default(resposta ~ (Método + Tipo))`



Interpretando os resultados

Na tabela de efeitos estimados, os baixos valores de p (0,040 e 0,000), para os dois fatores, indicam que o método de aplicação e o tipo de adesivo hot melt têm um impacto significativo sobre o tempo de colagem.

Observa-se o seguinte :

- o efeito do método de aplicação é negativo (**-0,13375**), porque o tempo de colagem é mais elevado, em média, quando a unidade experimental é **Manual** ao invés de **Spray**;
- o efeito do tipo de adesivo hot melt é positivo (0,47375), porque o tipo de **Adesivo B** teve uma média de tempo de adesão mais elevado do que o tipo de **Adesivo A**.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.67312	0.02931	159.463	< 2e-16 ***
Método1	-0.06688	0.02931	-2.282	0.04 *
Tipo1	0.23687	0.02931	8.083	0.000002 ***

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'
	0.1 ' '	1		

Residual standard error: 0.1172 on 13 degrees of freedom
 Multiple R-squared: 0.8444, Adjusted R-squared: 0.8205
 F-statistic: 35.27 on 2 and 13 DF, p-value: 0.000005601

```
Rcmdr> Anova(LinearModel.2, type="II")
Anova Table (Type II tests)
```

Response: resposta

	Sum Sq	Df	F value	Pr(>F)
Método	0.07156	1	5.2076	0.03997 *
Tipo	0.89776	1	65.3348	0.000001997 ***
Residuals	0.17863	13		

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'
	0.1 ' '	1		

```
Rcmdr> Confint(LinearModel.2, level=0.95)
Estimate 2.5 % 97.5 %
(Intercept) 4.673125 4.6098146 4.73643538
Método1 -0.066875 -0.1301854 -0.00356462
Tipo1 0.236875 0.1735646 0.30018538
```

```
Rcmdr> .bs.samples <- Boot(LinearModel.2, R=999, method="case")
Rcmdr+ plotBoot(.bs.samples)
```

Interpretando os resultados

A tabela ANOVA inclui um teste de falta de ajuste. A hipótese nula é de que este modelo ajusta-se adequadamente aos dados de resposta.

O valor-p (**0,315 > 0,10**) indica que não se deveria rejeitar esta hipótese. Neste caso, isto significa que não se cometeu um erro ao remover o termo de interação do modelo. Geralmente, a falta significativa de ajuste indica que termos importantes não foram incluídos no modelo.

A tabela de observações não usuais indica que a observação três é um ponto discrepante, pois seu resíduo supera a média zero em mais do que dois desvios padrão.

Os pontos discrepantes ocorrem frequentemente por acaso, porém, é habitualmente aconselhável procurar uma causa em potencial.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.67312	0.02931	159.463	< 2e-16 ***
Método1	-0.06688	0.02931	-2.282	0.04 *
Tipo1	0.23687	0.02931	8.083	0.000002 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Residual standard error: 0.1172 on 13 degrees of freedom
 Multiple R-squared: 0.8444, Adjusted R-squared: 0.8205
 F-statistic: 35.27 on 2 and 13 DF, p-value: 0.000005601

```
Rcmdr> Anova(LinearModel.2, type="II")
Anova Table (Type II tests)
```

Response: resposta

	Sum Sq	Df	F value	Pr(>F)
Método	0.07156	1	5.2076	0.03997 *
Tipo	0.89776	1	65.3348	0.000001997 ***
Residuals	0.17863	13		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

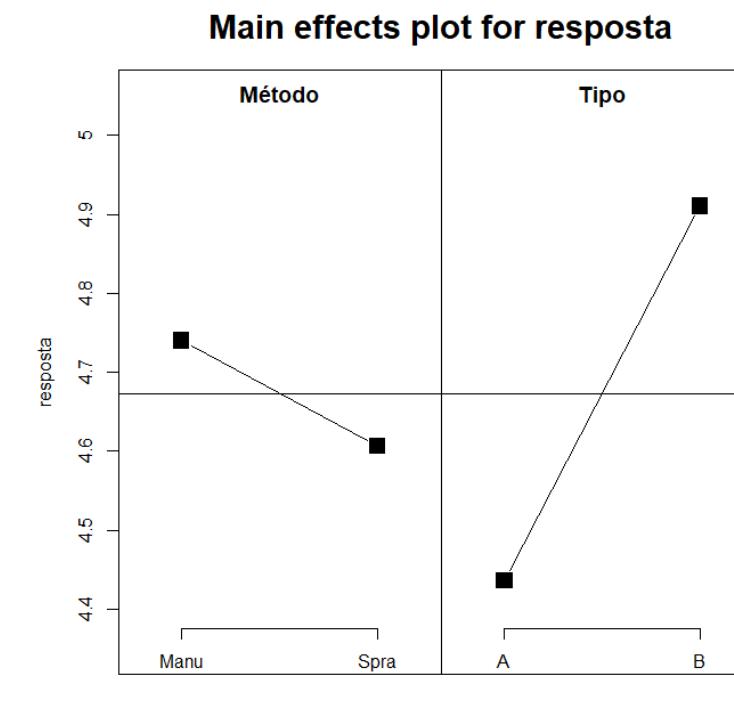
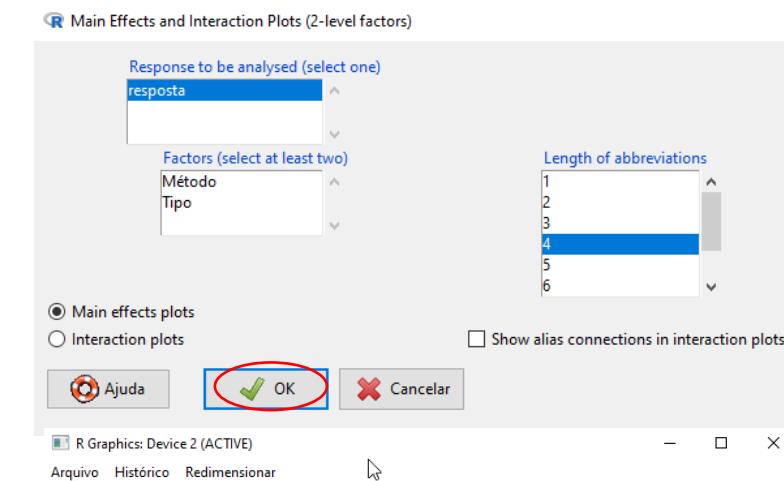
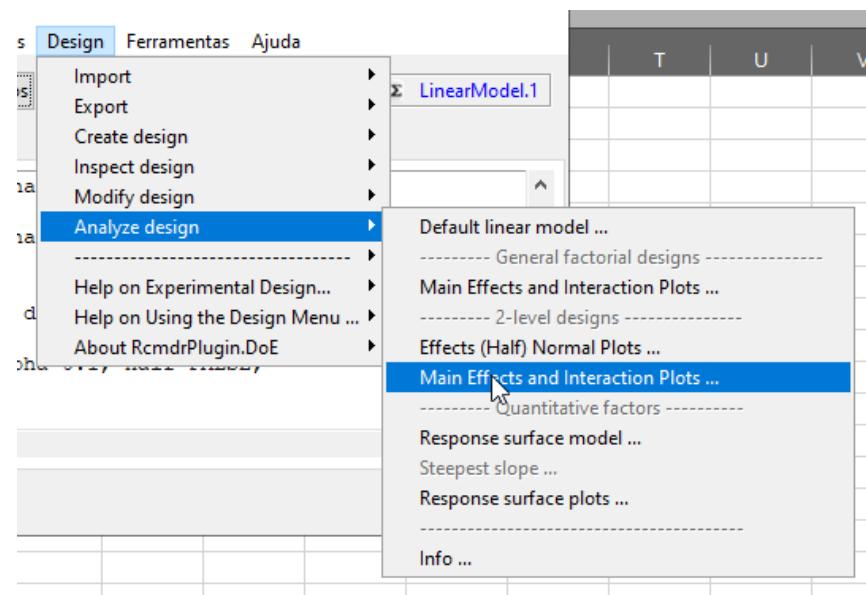
	Estimate	2.5 %	97.5 %
(Intercept)	4.673125	4.6098146	4.73643538
Método1	-0.066875	-0.1301854	-0.00356462
Tipo1	0.236875	0.1735646	0.30018538

```
Rcmdr> .bs.samples <- Boot(LinearModel.2, R=999, method="case")
Rcmdr+ plotBoot(.bs.samples)
```

Experimentos Fatoriais Completos

Efeitos principais

- O método de aplicação (Fator A) tem um efeito negativo, pois o tempo médio de adesão para aplicação com **spray** é mais baixo que para aplicação **manual** (lembre-se de que manual é a configuração baixa e spray é a configuração alta do fator método de aplicação);
- O tipo de adesivo hot melt (Fator B) tem um efeito positivo, pois o tipo de adesivo **B** (importado) tem um tempo médio de adesão mais elevado do que o tipo de adesivo **A** (nacional).

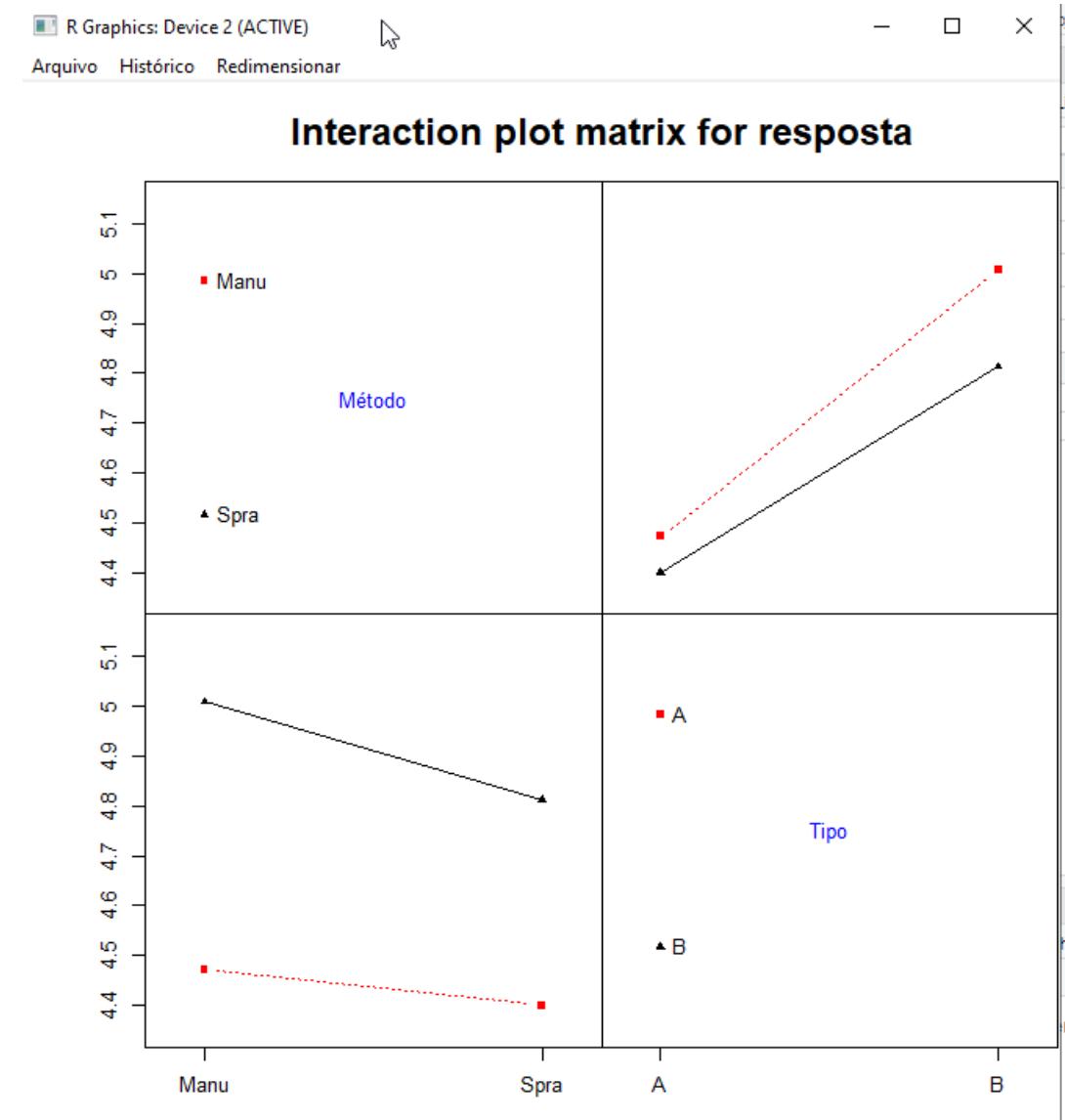
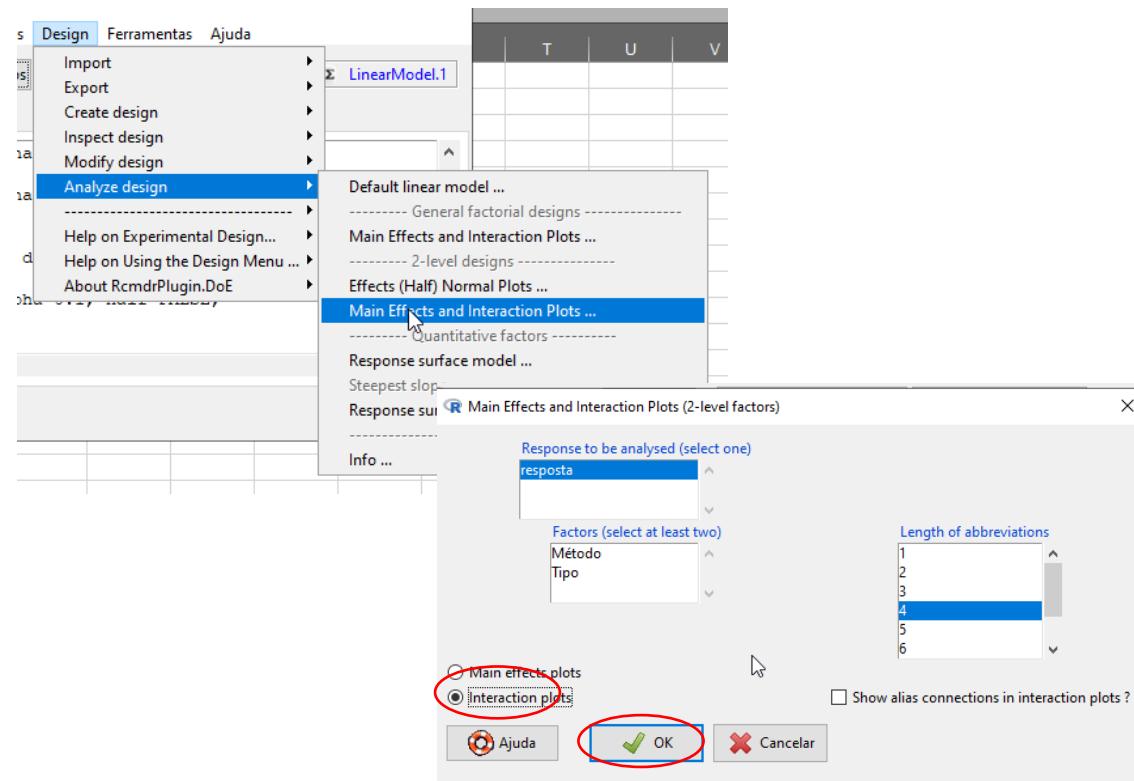


Interpretando os resultados

Gráfico da interação

O gráfico da interação mostra que, independente do tipo de adesivo, o método de aplicação com **Spray** resulta em uma adesão mais rápida.

Sabe-se que a interação não é importante pelo fato de que as duas linhas neste gráfico são quase paralelas. Confirmou-se que a interação não foi significativa pelo elevado valor-p associado com o efeito da interação a partir do modelo completo.



Considerações finais

Conclusões práticas

O tempo de colagem é minimizado quando aplicamos o adesivo do tipo A (Nacional) com o método de aplicação com Spray. Contudo, deve-se considerar o custo de cada método de aplicação e tipo de adesivo hot melt. Pode-se optar por utilizar configurações de fatores menos “ótimas”, se a melhoria não for suficiente para justificar um aumento substancial nos custos de produção.

Considerações estatísticas

Para ...

Use ...

identificar termos importantes no modelo

um gráfico de Pareto ou um gráfico de Probabilidade normal dos efeitos

obter informações sobre a constante e efeito de blocos

a tabela de coeficientes e efeitos

analisar a significância de grupos de efeitos

a tabela ANOVA

considerar o efeito da remoção simultânea de dois ou mais termos

um teste de falta de ajuste

avaliar a normalidade, independência e variância de igualdade dos resíduos

gráfico de resíduos

visualizar efeitos principais e interações two-way

gráfico de efeitos principais e de interação

mostrar resultados médios em todas as combinações de tratamento

gráficos de cubo



13

Experimentos Fatoriais Fracionados

Nível: Mestre

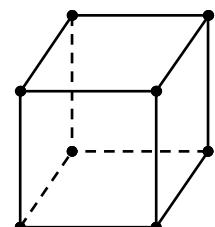
Experimentos Fatoriais Fracionados

O que são experimentos fatoriais fracionados ?

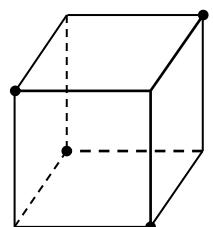
Para a maioria dos experimentadores, a quantidade de recursos disponíveis (tempo, dinheiro) é limitado. À medida que o número de fatores em um experimento fatorial 2^k aumenta, o número de ensaios necessários para efetuar uma réplica completa aumenta rapidamente. Em muitos casos, o experimento fatorial completo torna-se muito caro ou muito difícil de executar.

Os experimentos fatoriais fracionados utilizam um subconjunto (ou fração) de um experimento fatorial completo para obter informações sobre os efeitos principais e as interações de menor ordem.

Fatorial Completo



1/2 Fração



Quando usar experimentos fatoriais fracionados ?

Um experimento 2^7 completo requer 128 ensaios por réplica. O modelo completo contém 7 efeitos principais e 21 interações de dois fatores. O resto dos termos são interações de ordem mais alta.

Caso as interações de ordem mais alta possam considerar-se desprezíveis, podem-se obter informações suficientes sobre os efeitos principais e as interações duplas executando apenas uma fração do experimento fatorial.

Por que usar experimentos fatoriais fracionados ?

São utilizados para responder a questões tais como :

- quais fatores têm grande efeito na resposta ?
- quais fatores têm um impacto desprezível na resposta ?
- quais as interações de menor ordem que são importantes para explicar a variação na resposta ?

Por exemplo :

- quais são os possíveis fatores que levam aos vazamentos em latas de aerosol ?
- existe alguma interação dupla entre a temperatura de molde, a velocidade de parafuso, o tempo de retenção, o tempo de ciclo e a temperatura de retenção que seja importante para explicar o encolhimento excessivo no processo de moldagem ?

Experimentação Sequencial

Criando um experimento de $\frac{1}{2}$ fração

Notação

Observa-se que, na subcaixa de diálogo Designs (Experimentos), o experimento também é chamado 2^{5-1} . Na verdade, este é um experimento $2^5 \times (\frac{1}{2})$. Uma vez que $\frac{1}{2}$ é o mesmo que 2^{-1} , então $2^5 \times 2^{-1} = 2^{5-1}$. Esta notação indica que o experimento acomodará cinco fatores, cada um em dois níveis, mas que apenas $2^{5-1} = 2^4 = 16$ ensaios serão empregados.

Requested information
You may close this window for continuing work with the design dialogue.

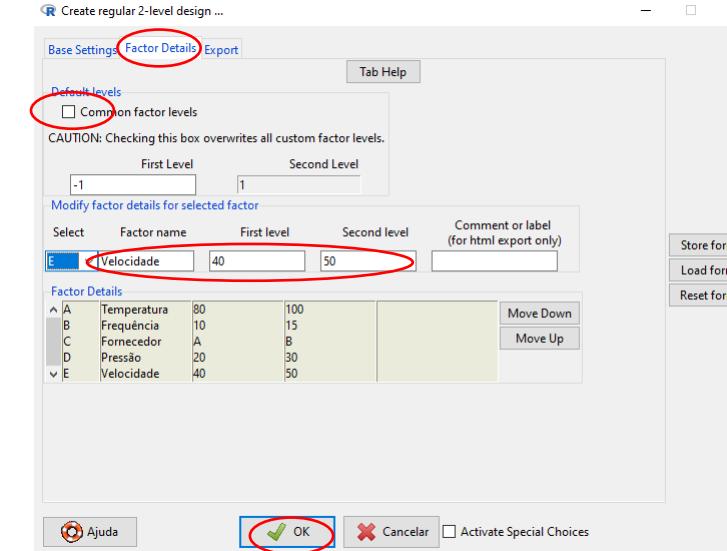
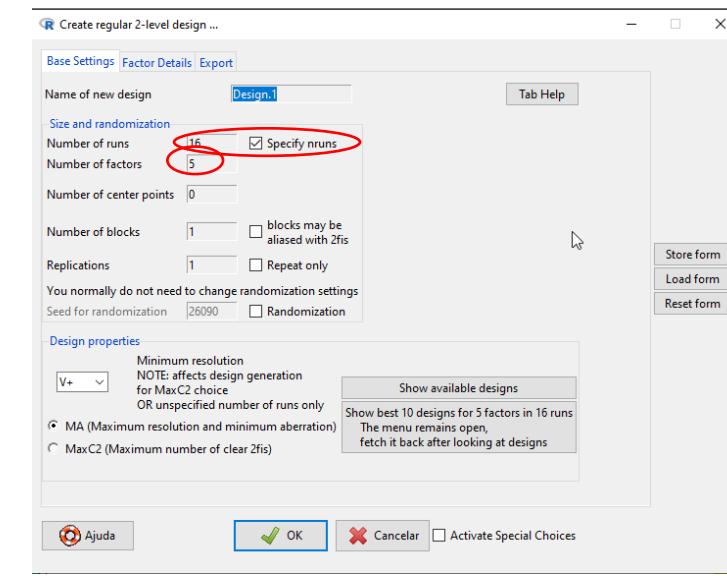
	8	16	32	64	128	256	512	1024	2048	4096
3	full									
4	IV	full								
5	III	V	full							
6	III	IV	VI	full						
7	III	IV	IV	VII	VIII					
8	III	IV	V	V	VIII	full				
9	III	IV	IV	VI	VII	X	full			
10	III	IV	IV	V	VII	XI	full			
11	III	IV	IV	V	VII	XI	full			
12	III	IV	IV	V	VII	VIII	XII	full		
13	III	IV	IV	V	VII	VIII	VIII	XIII	full	
14	III	IV	IV	V	VII	VIII	VIII	XIV	full	
15	III	IV	IV	V	VII	VIII	VIII	VIII	XV	full
16	IV	IV	IV	V	VII	VIII	VIII	VIII	VIII	XVI
17	III	IV	IV	V	VII	VIII	VIII	VIII	VIII	XVII
18	III	IV	IV	IV	VII	VIII	VIII	VIII	VIII	XVIII
19	III	IV	IV	IV	V	VII	VIII	VIII	VIII	XIX
20		III	IV	IV	IV	V	VII	VIII	VIII	XX
21		III	IV	IV	IV	V	VII	VIII	VIII	XI
22		III	IV	IV	IV	V	VII	VIII	VIII	XII
23		III	IV	IV	IV	V	VII	VIII	VIII	XIII
24		III	IV	IV	IV	V	VII	VIII	VIII	XIV

number of factors

Resolution III up to 31 63 127 factors.
Resolution IV up to 32 64 80 160 factors.
Resolution V up to number of factors: 33 47 65
Resolution VI up to number of factors: 24 34 48
First design is MA up to number of factors: 31 63 127 36 29 28 32 26

Additional irregular resolution V designs from general orthogonal arrays menu:
128 runs (up to 15 factors), 256 runs (up to 19 factors), 2048 runs (up to 63 factors)

Criar Experimentos Fatoriais



Interpretando os resultados

Uma vez que o fator **E (Velocidade)** tenha sido colocado na coluna da interação de quarta ordem, qualquer efeito adivinho de **E** será confundido com o efeito da interação **ABCD**. O pressuposto é que o efeito da interação de quarta ordem seria desprezível. Portanto, qualquer efeito significativo seria atribuído ao fator **E** (velocidade, em m/s).

```
Rcmdr> print( Design.1 )
  Temperatura Frequência Fornecedor Pressão Velocidade resposta2
1       80        10          A      20       50       56
2      100        10          A      20       40       53
3       80        15          A      20       40       63
4      100        15          A      20       50       65
5       80        10          B      20       40       53
6      100        10          B      20       50       55
7       80        15          B      20       50       67
8      100        15          B      20       40       61
9       80        10          A      30       40       69
10      100        10          A      30       50       45
11      80        15          A      30       50       78
12      100        15          A      30       40       93
13      80        10          B      30       50       49
14      100        10          B      30       40       60
15      80        15          B      30       40       95
16      100        15          B      30       50       82
class=design, type= FrF2
```

Interpretando os resultados

A tabela de experimento relaciona o seguinte :

Factors – Existem 5 fatores em seu experimento

Base Design – Este é o experimento base sem réplicas. Este experimento base tem 5 fatores e 16 ensaios

Runs – Este experimento tem 16 ensaios

Replicates – Tem-se uma única réplica em cada ponto do experimento

Blocks – Este experimento não possui blocos. Em outras palavras, todos os pontos do experimento estão contidos em um único bloco

Center pts (total) – Não existem pontos centrais neste experimento

Resolution – Em um experimento de resolução V, os efeitos principais são confundidos com interações de quarta ordem ($1+4=5$), e as interações duplas são confundidas com as interações de terceira ordem ($2+3=5$)

Fraction – Está sendo utilizado $\frac{1}{2}$ dos pontos experimentais que seriam incluídos em um experimento fatorial completo.

```
Rcmdr> summary( Design.1 , brief = TRUE)
design was generated with RcmdrPlugin.DoE

Experimental design of type FrF2
16 runs

Factor settings (scale ends):
  Temperatura Frequência Fornecedor Pressão Velocidade
1       80        10          A      20       40
2      100        15          B      30       50

Responses:
[1] resposta2

Design generating information:
$legend
[1] A=Temperatura B=Frequência C=Fornecedor D=Pressão   E=Velocidade
```

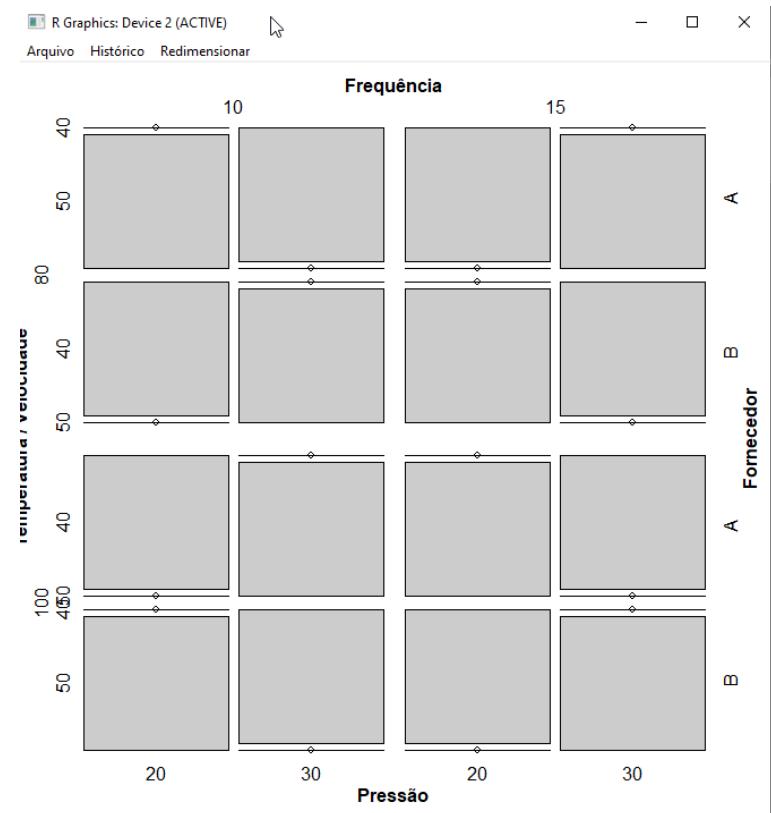
Interpretando os resultados

Geradores de experimento

O fator **E** tem o mesmo padrão que **ABCD**. Isto é mostrado na janela **Session** através da seguinte notação :

Designs Generators : E = ABCD

```
$generators
[1] E=ABCD
```



Interpretando os resultados

Identidade e confundimento

A janela **Session** também exibe a estrutura de confundimento do experimento.

A chave para o padrão de confundimento para o experimento inteiro é dado na *identidade* :

I + ABCDE

Para determinar o confundimento para um termo, multiplica-se o termo de interesse pela identidade e, em seguida, elimina-se quaisquer termos quadráticos.

Por exemplo, para saber com que efeito o termo **BC** é confundido deve-se fazer :

$$(BC)*(I+ABCDE) = BC + AB^2C^2DE = BC + ADE$$

Portanto, **BC** e **ADE** são confundidos um com o outro. Isso significa que qualquer efeito atribuído à interação **B*C** não pode ser separado de qualquer efeito causado pela interação **A*D*E**.

Interpretando os resultados

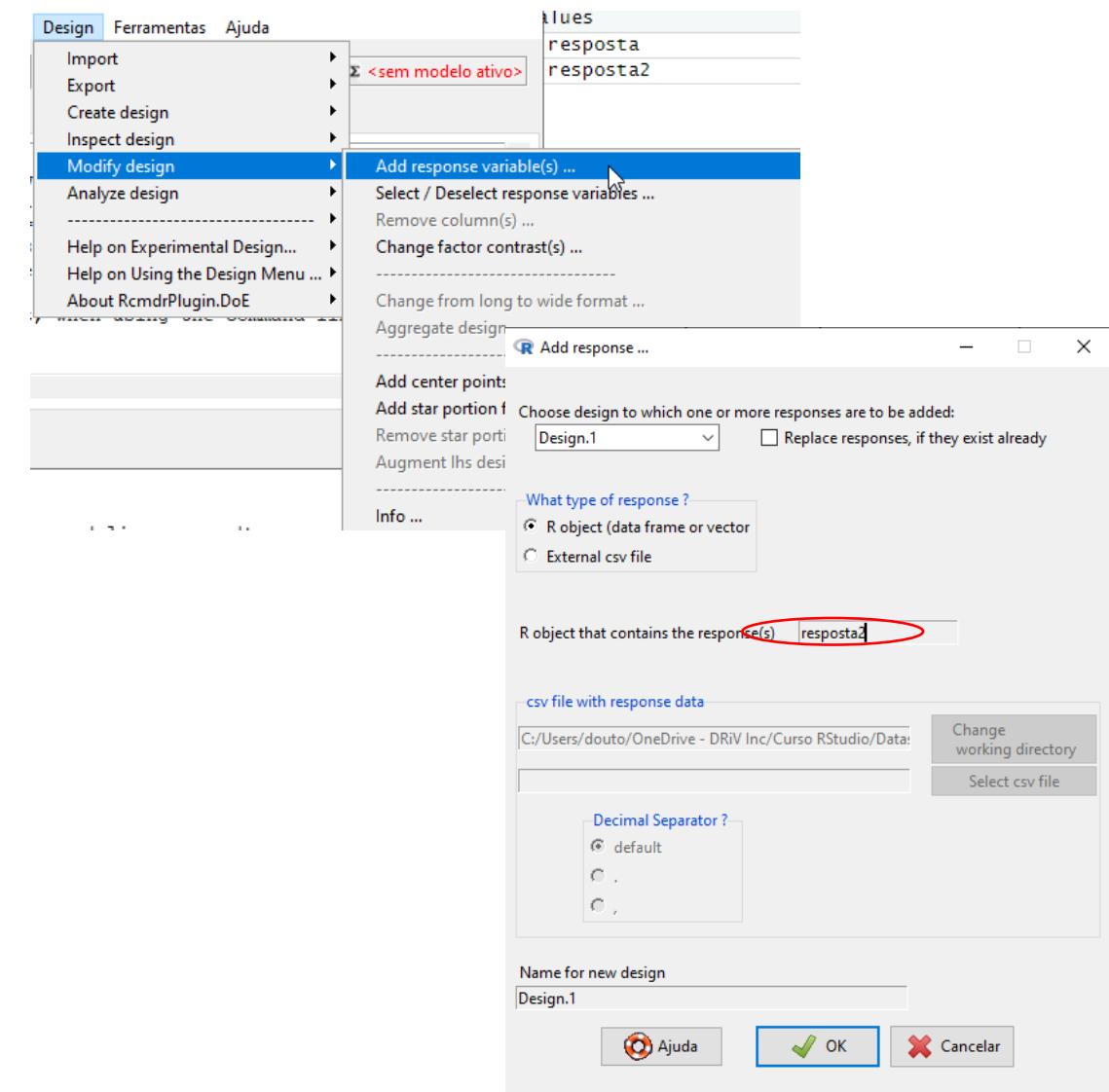
É conveniente certificar-se de que a janela Data corresponde ao que se está vendo aqui antes de continuar com a análise.

```
Rcmdr> print( Design.1 )
    Temperatura Frequência Fornecedor Pressão Velocidade resposta2
 1          80       10           A      20       50      56
 2         100       10           A      20       40      53
 3          80       15           A      20       40      63
 4         100       15           A      20       50      65
 5          80       10           B      20       40      53
 6         100       10           B      20       50      55
 7          80       15           B      20       50      67
 8         100       15           B      20       40      61
 9          80       10           A      30       40      69
10         100       10           A      30       50      45
11          80       15           A      30       50      78
12         100       15           A      30       40      93
13          80       10           B      30       50      49
14         100       10           B      30       40      60
15          80       15           B      30       40      95
16         100       15           B      30       50      82
class=design, type= FrF2
```

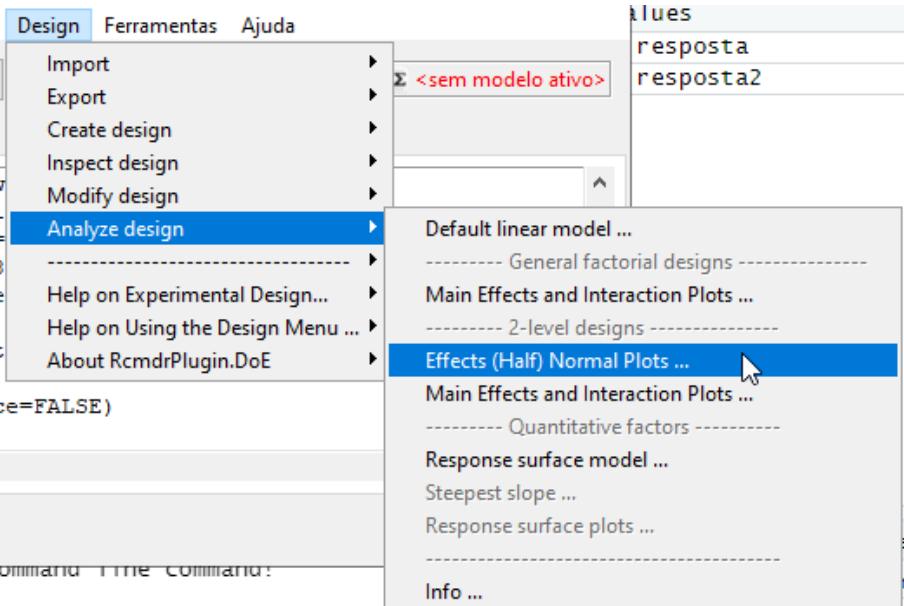
Analizando um experimento fatorial fracionado

Este experimento é de Resolução V e não tem réplicas. Portanto, os únicos efeitos que se podem estimar, neste estudo, São os cinco efeitos principais e as dez interações de dois fatores. Deve-se presumir que todas as interações de três, quatro e cinco fatores são desprezíveis.

Analizar um Experimento Fatorial



Analisar um experimento factorial:



Design Ferramentas Ajuda

- Import
- Export
- Create design
- Inspect design
- Modify design
- Analyze design
 - Default linear model ...
 - General factorial designs
 - Main Effects and Interaction Plots ...
 - 2-level designs
 - Help on Experimental Design...
 - Help on Using the Design Menu ...
 - About RcmdrPlugin.DoE

:e=FALSE)

half-normal scores

Responses

resposta
resposta2

absolute effects

Effects plots for 2-level factors

Response to be analysed (select one)
resposta2

Half normal plot

Label effects with codes instead of names ?

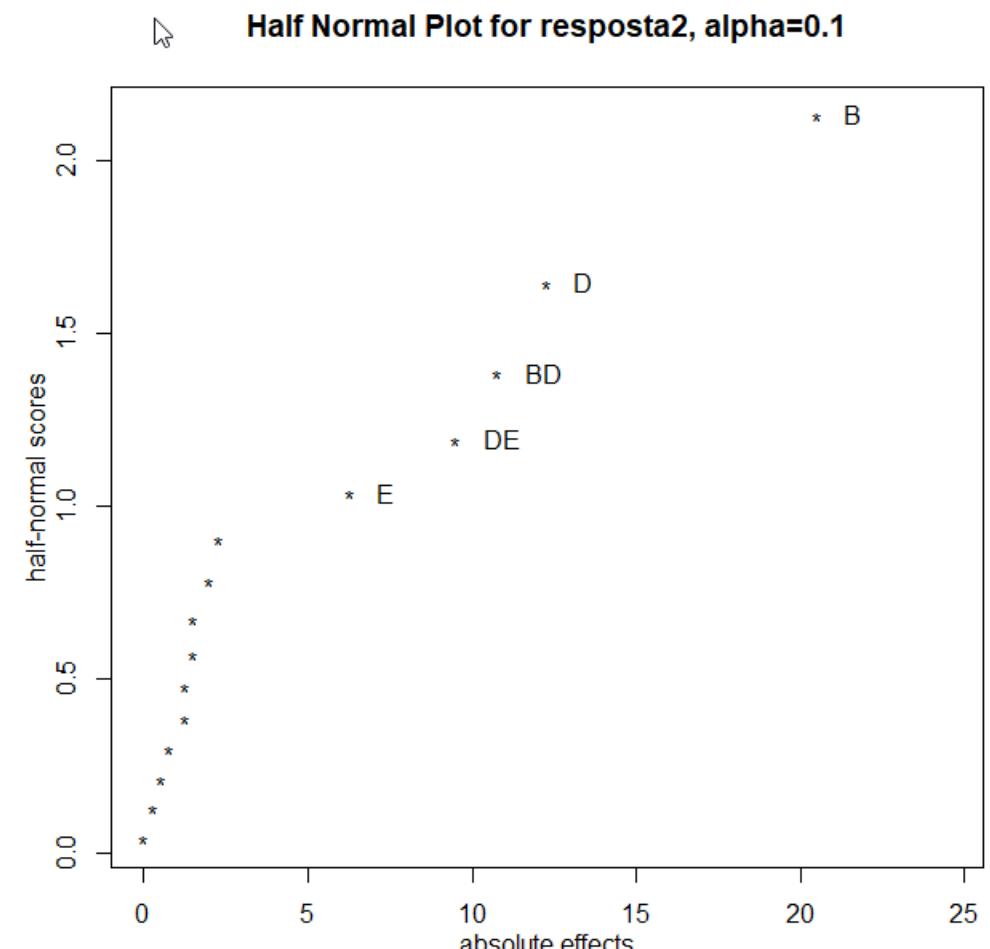
Enter significance level for labelling:
0.1

Label significant effects only ?

Ajuda OK Cancelar

Desta vez, vamos
Deixar a meia-
Normal ativada:

A grande diferença é que ele joga todos os fatores relevantes para a direita.



A = Temperatura , B = Frequência , C = Fornecedor , D = Pressão , E = Velocidade

Interpretando os resultados

Gráfico de probabilidade normal

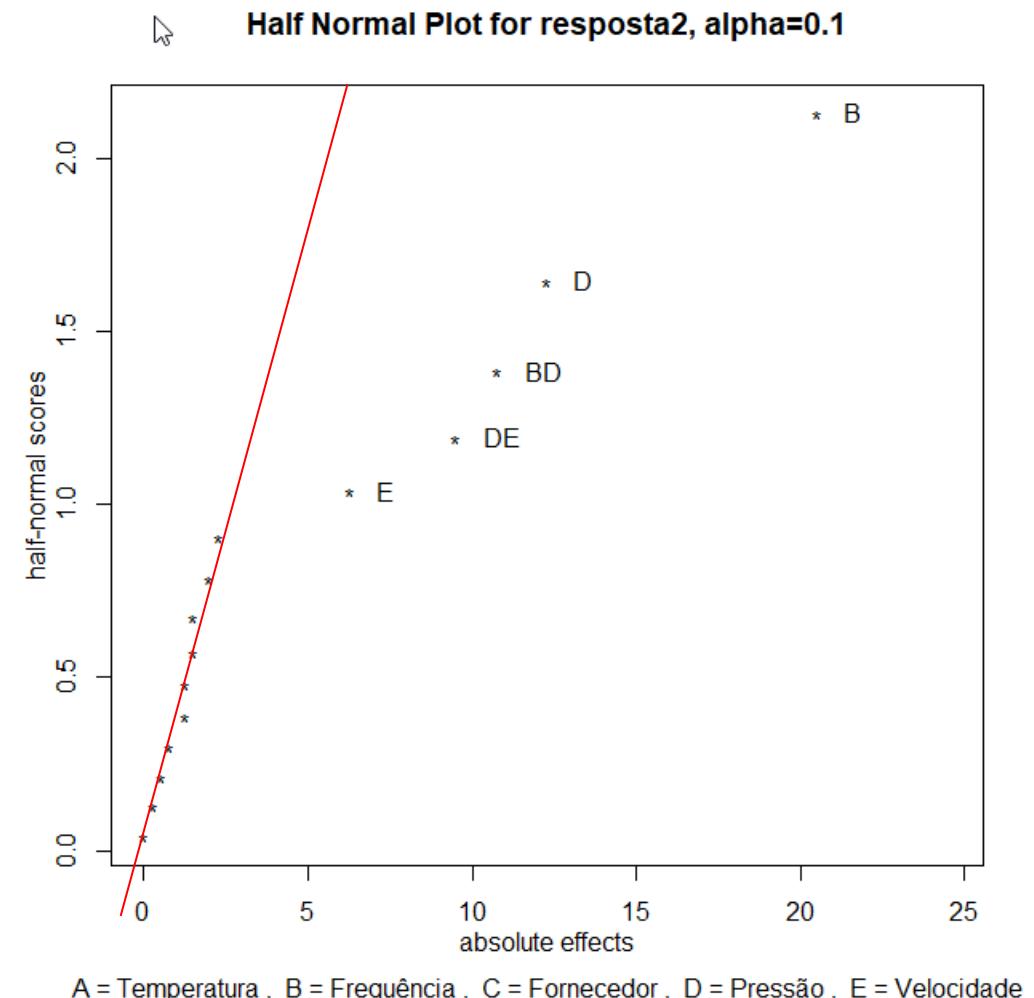
Se os efeitos dos fatores fossem zero e a variação devida apenas ao acaso, todos os pontos cairiam em uma linha quase reta.

Aqueles efeitos distantes da linha são considerados significativos, pois seus valores são de uma magnitude tal que a probabilidade deles ocorrerem por acaso é baixa.

Note que no caso ao lado apenas aparece com título os fatores significativos

A linha vermelha foi adicionada manualmente, o Rstudio não mostra ela, mas na verdade nem é necessário.

Logo, os fatores B, D e E são significativos, bem como as interações BD e DE. Todas as demais não!



Interpretando os resultados

Vamos então gerar o modelo linear, mas, para não perder tempo, como já vimos que alguns fatores não importam, vamos remove-los de cara

The screenshot shows the Rcmdr interface. The 'Analyze design' menu is open, and 'Default linear model ...' is selected. In the main window, the 'Linear model for experimental designs' dialog box is displayed. The formula input field contains the text: 'resposta2 ~ Frequencia + Pressao + Velocidade + Frequencia*Pressao + Pressao*Velocidade'. The word 'Frequencia' is highlighted with a red oval.

Análise do Experimento Fatorial

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 65.2500 0.6626 98.473 2.86e-16 ***
Frequencial 10.2500 0.6626 15.469 2.60e-08 ***
Pressão1 6.1250 0.6626 9.244 3.25e-06 ***
Velocidade1 -3.1250 0.6626 -4.716 0.000821 ***
Frequencial:Pressão1 5.3750 0.6626 8.112 1.04e-05 ***
Pressão1:Velocidade1 -4.7500 0.6626 -7.169 3.04e-05 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.65 on 10 degrees of freedom
Multiple R-squared: 0.9789, Adjusted R-squared: 0.9684
F-statistic: 92.83 on 5 and 10 DF, p-value: 0.00000004766
```

```
Rcmdr> Anova(LinearModel.4, type="II")
Anova Table (Type II tests)
```

```
Response: resposta2
Sum Sq Df F value    Pr(>F)
Frequência 1681.00 1 239.288 0.000000026 ***
Pressão    600.25 1 85.445 0.000003253 ***
Velocidade 156.25 1 22.242 0.0008212 ***
Frequência:Pressão 462.25 1 65.801 0.000010425 ***
Pressão:Velocidade 361.00 1 51.388 0.000030372 ***
Residuals   70.25 10
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tabela da Análise de Variância

A tabela de análise de variância, na janela **Session**, mostra que os efeitos principais e interações duplas são significativos (ambos os valores p são < 0,01).

Interpretando os resultados

Coeficientes estimados

O MINITAB exibe os coeficientes não codificados na base da janela Session. O modelo em unidades não codificadas (reais) é :

$$\text{Produção} = 65,25 + 10,25 \cdot \text{Frequência} + \\ 6,125 \cdot \text{Pressão} - 3,125 \cdot \text{Velocidade} + \\ 5,375 \cdot \text{Frequência} \cdot \text{Pressão} - \\ 4,75 \cdot \text{Pressão} \cdot \text{Velocidade}$$

Esta equação pode ser usada com unidades reais. Para usar esta equação para prever a quantidade de Produção para configurações específicas de Frequência, Pressão e Velocidade utiliza-se valores codificados (-1 e +1).

```

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 65.2500 0.6626 98.473 2.86e-16 ***
Frequêncial 10.2500 0.6626 15.469 2.60e-08 ***
Pressão1 6.1250 0.6626 9.244 3.25e-06 ***
Velocidade1 -3.1250 0.6626 -4.716 0.000821 ***
Frequêncial:Pressão1 5.3750 0.6626 8.112 1.04e-05 ***
Pressão1:Velocidade1 -4.7500 0.6626 -7.169 3.04e-05 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.65 on 10 degrees of freedom
Multiple R-squared: 0.9789, Adjusted R-squared: 0.9684
F-statistic: 92.83 on 5 and 10 DF, p-value: 0.00000004766

```

```

Rcmdr> Anova(LinearModel.4, type="II")
Anova Table (Type II tests)

Response: resposta2
Sum Sq Df F value    Pr(>F)
Frequência 1681.00 1 239.288 0.000000026 ***
Pressão    600.25 1 85.445 0.000003253 ***
Velocidade 156.25 1 22.242 0.0008212 ***
Frequência:Pressão 462.25 1 65.801 0.000010425 ***
Pressão:Velocidade 361.00 1 51.388 0.000030372 ***
Residuals   70.25 10
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

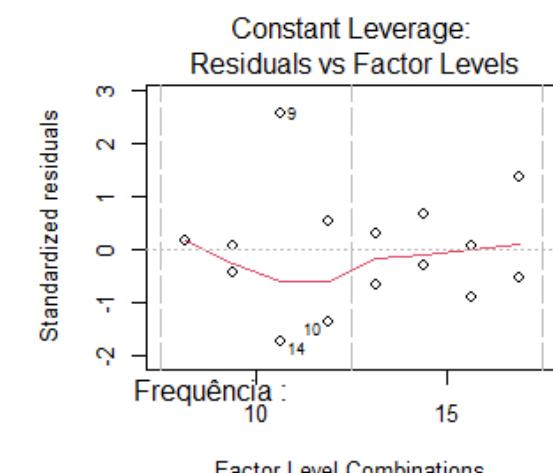
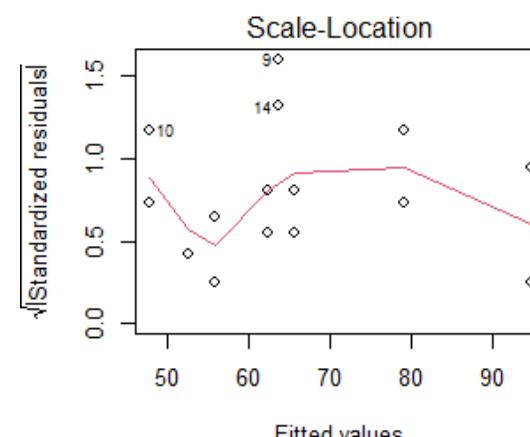
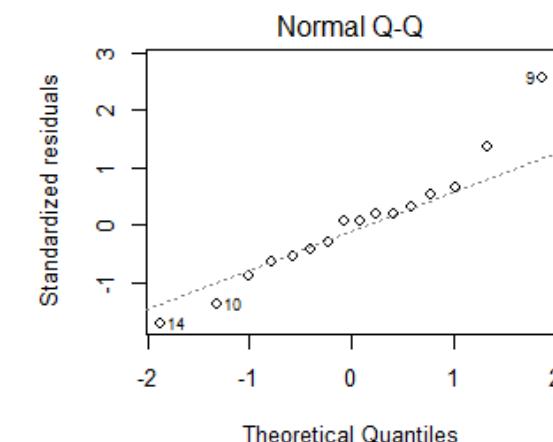
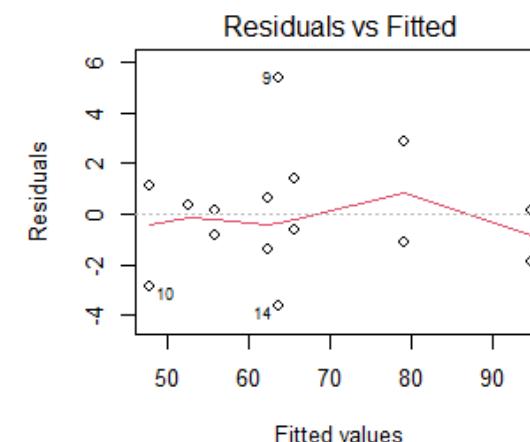
Interpretando os resultados

Utiliza-se o gráfico de probabilidade normal dos resíduos para verificar se estes não se desviam de forma substancial de uma distribuição normal.

- se os resíduos são provenientes de uma distribuição normal, os pontos seguem, a grosso modo, uma linha reta;
- se os resíduos não são provenientes de uma distribuição normal, os pontos não seguem uma linha reta.

Para os dados de produtividade, a linha apresentada no **Normal Probability Plot of the Residuals** (Gráfico de Probabilidade Normal dos Resíduos) está próximo o suficiente de uma linha reta para se presumir que os erros sejam normais.

`default(resposta2 ~ Frequência + Pressão + Velocidade + Frequência * Pre`



Interpretando os resultados

Utiliza-se o gráfico dos resíduos versus os valores ajustados para verificar se os seguintes pressupostos são satisfeitos :

- variância constante ao longo de todos os níveis de fatores
- não existe nenhum ponto discrepante dos dados

A tabela abaixo resume os padrões típicos possíveis de serem encontrados. A ocorrência desses padrões em um gráfico indica que um dos pressupostos foi violado.

Este padrão ...

dispersão em漏il ou desigual dos resíduos ao longo de diferentes valores ajustados

um ponto que se encontra muito longe de zero

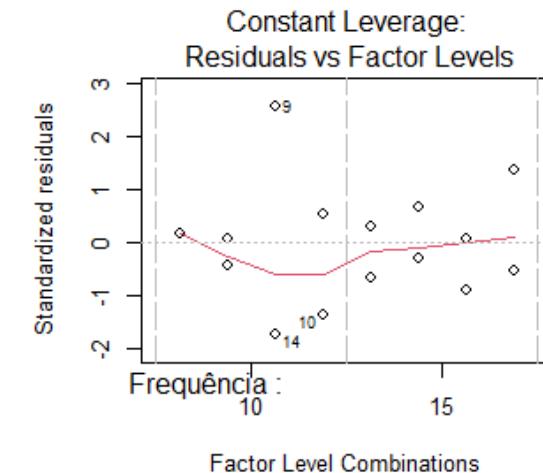
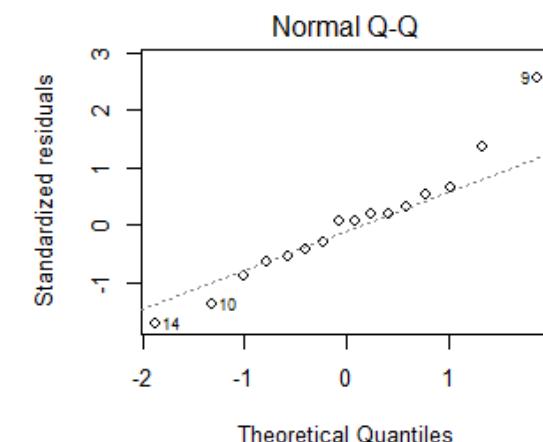
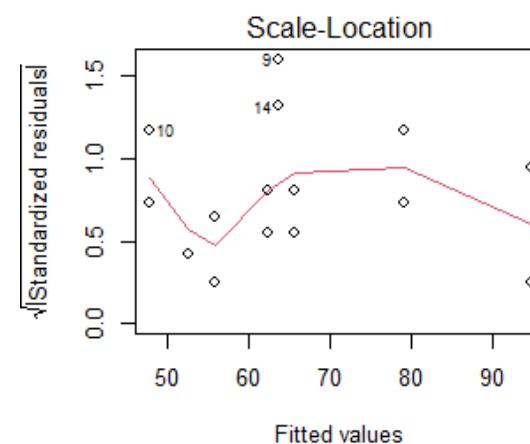
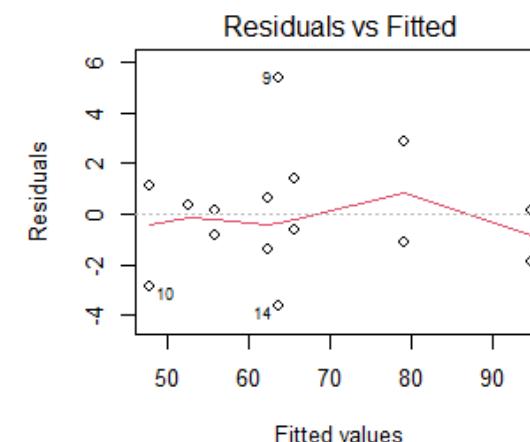
Para os dados de produtividade, o gráfico de **Residuals Versus the Fitted Values** (Gráfico dos Resíduos Versus os Valores Ajustados) mostra uma dispersão aleatória, indicativa de uma variância constante dos resíduos.

Indica ...

a variância de seus resíduos não é constante

um ponto discrepante

default(resposta2 ~ Frequência + Pressão + Velocidade + Frequência * Pre



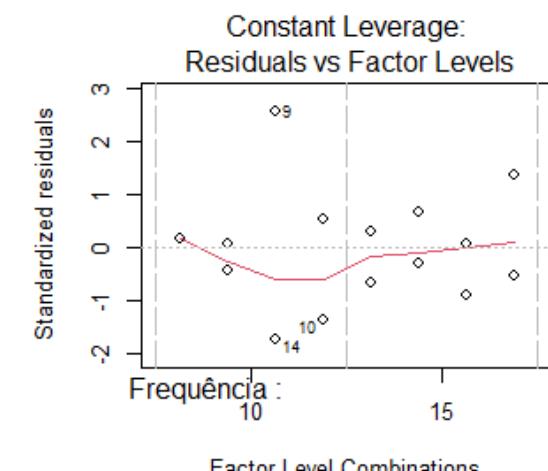
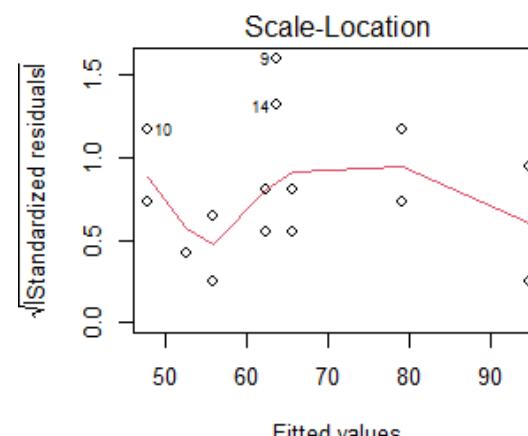
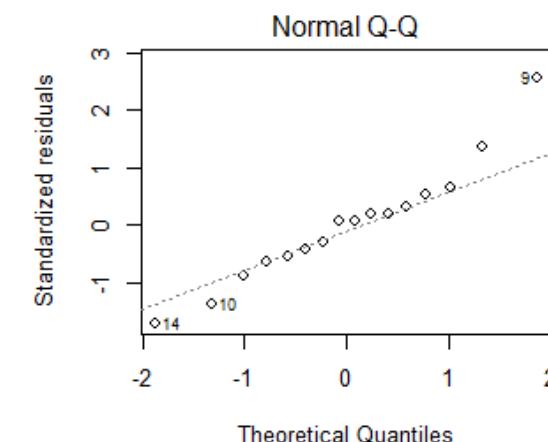
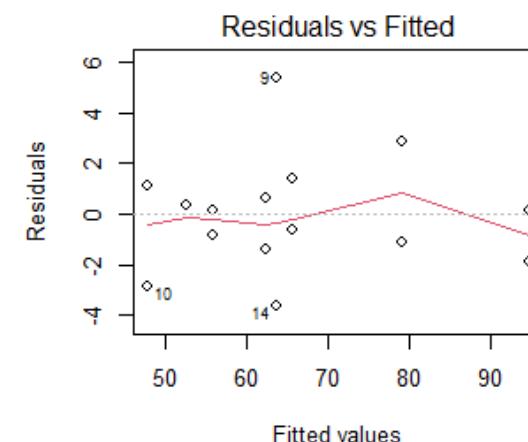
Interpretando os resultados

Utiliza-se o gráfico dos resíduos versus a ordem dos dados para verificar se os resíduos são independentes.

- se houver efeito devido à ordem de coleta dos dados, os resíduos não estarão distribuídos aleatoriamente em torno de zero e um padrão no gráfico será detectado;
- se não houver efeito devido à ordem de coleta dos dados, os resíduos estarão distribuídos aleatoriamente em torno de zero.

Para os dados de produtividade, o gráfico de **Residuals Versus the Order of the Data** (gráfico de resíduos versus ordem dos dados), não mostra qualquer sinal de tendência nos erros devido à ordem dos ensaios experimentais.

`default(resposta2 ~ Frequência + Pressão + Velocidade + Frequência * Pre`



Encontrando as melhores configurações

O gráfico dos fatores pode ser utilizado para encontrar as melhores configurações para a Frequência, Pressão e Velocidade, de modo a maximizar a produtividade da linha de montagem de lateral de porta.

Use
Ctrl+
clique
para
pintar de
azul

Design Ferramentas Ajuda

- Import
- Export
- Create design
- Inspect design
- Modify design
- Analyze design**
-
- Help on Experimental Design...
- Help on Using the Design Menu ...
- About RcmdrPlugin.DoE

LinearModel.4

- LinearModel.4
- oldpar
- values
- resposta
- resposta2

- Default linear model ...
- General factorial designs -----
- Main Effects and Interaction Plots ...
- 2-level designs -----
- Effects (Half) Normal Plots ...
- Main Effects and Interaction Plots ...**
- Quantitative factors -----
- Response surface model ...
- Steepest slope ...

Main Effects and Interaction Plots (2-level factors)

Response to be analysed (select one)
resposta2

Factors (select at least two)
Frequência

Length of abbreviations
1
2
3
4
5
6

Main effects plots

Show alias connections in interaction plots ?

Ajuda OK Cancelar

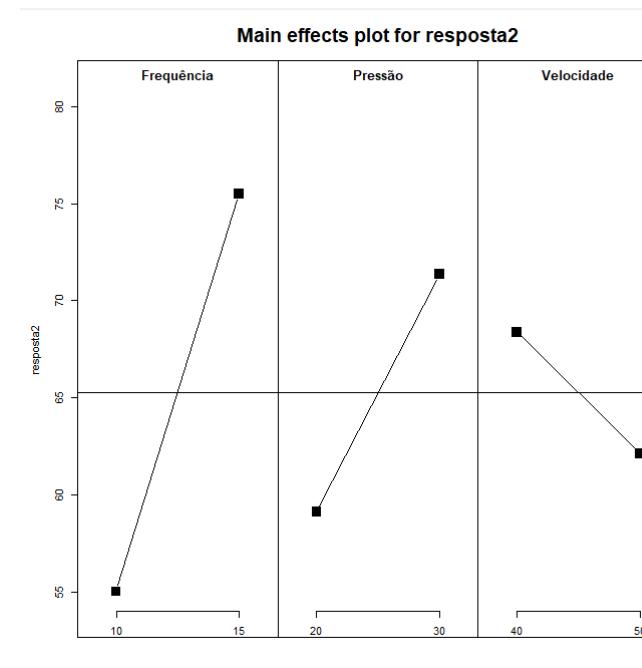
Interpretando os resultados

Gráfico dos efeitos principais

O gráfico de efeitos principais mostra que a melhor configuração para aumento da produtividade da linha de lateral de porta é a seguinte :

- frequência 15 (+1), Pressão 30 (+1) e Velocidade 40 (-1).

Estes efeitos principais, contudo, podem ser enganadores na presença de interações significativas, por isso é importante analisar também o gráfico das interações e o gráfico de cubo.

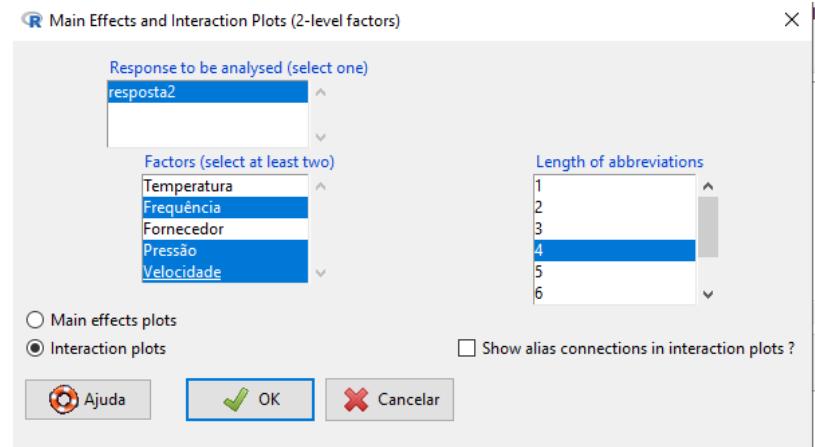


Interpretando os resultados

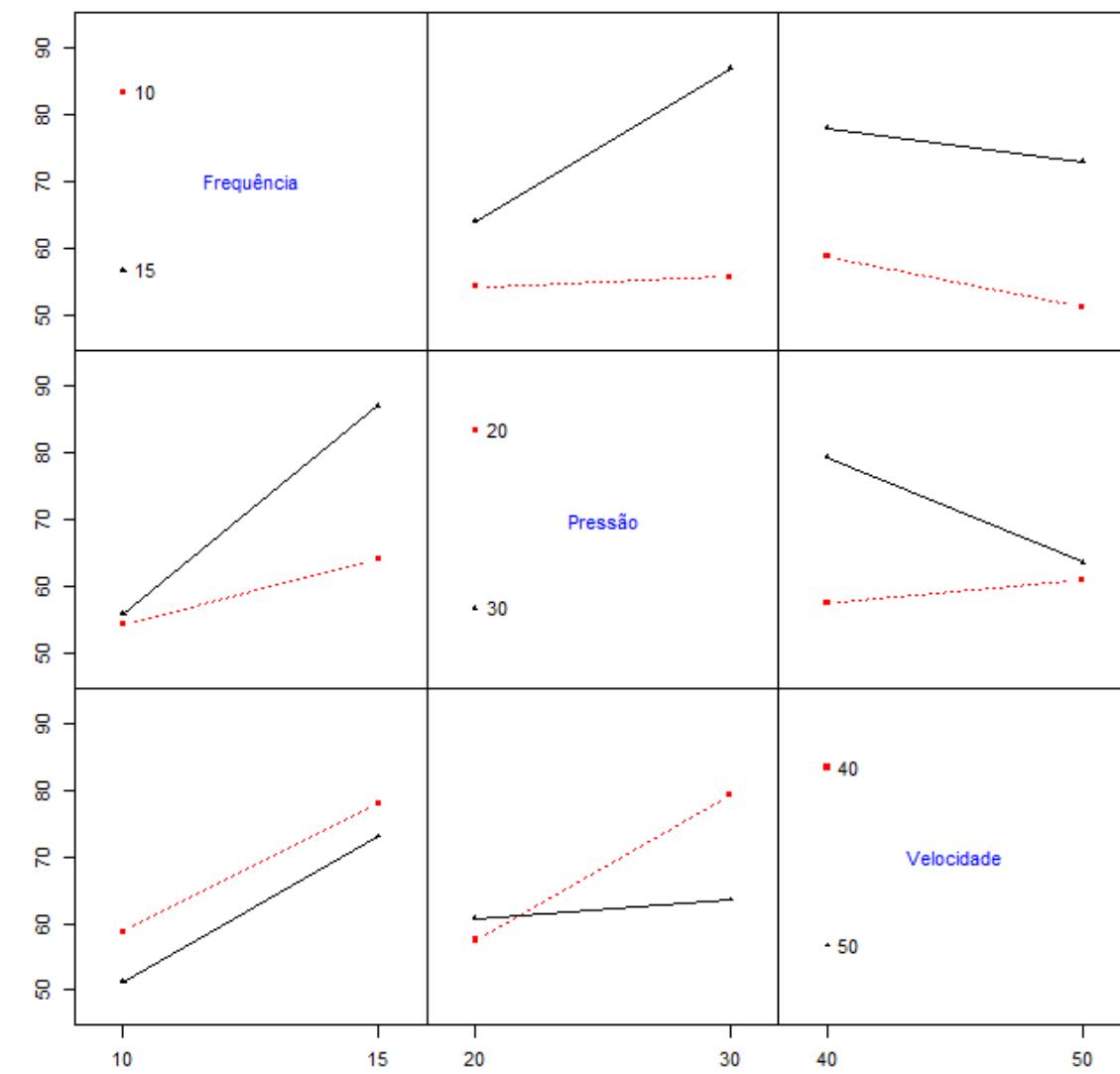
Gráfico de interações

O gráfico da interações mostra o seguinte :

- Interação entre Frequência e Pressão – A pressão de 30 (+1) sempre apresenta melhores resultados de produtividade. No caso da frequência, a melhor configuração seria 15 (+1), já que a mesma é sempre maior que a frequência 10 (-1);
- Interação entre Frequência e Velocidade – A velocidade de 40 (-1) sempre apresenta melhores resultados de produtividade. No caso da frequência, a configuração 15 (+1) ainda é a melhor opção;
- Interação Pressão e Velocidade – A velocidade de 40 (-1), continua apresentando o melhor resultado de produtividade. No caso da pressão, a configuração de 30 (+1) seria a melhor opção.



Interaction plot matrix for resposta2



Considerações finais

Conclusões práticas

Pode-se concluir que para maximizar a produtividade na reação :

- a taxa de velocidade, frequência e pressão apresentam impacto significativo;
- a frequência 15 Hz, a pressão 30 psi e a velocidade 40 m/s maximizam a produtividade da linha de montagem da lateral de porta.

Considerações estatísticas

Iniciou-se o trabalho com $\frac{1}{2}$ fração, um experimento factorial fracionado com 16 ensaios para 5 fatores. Contudo, 2 dos fatores não foram significativos, portanto os 16 ensaios tornaram-se um experimento factorial completo com 2 réplicas para os 3 fatores significativos - a frequência, a pressão e a velocidade.

Tem-se ainda dinheiro no orçamento para ensaios adicionais. Pode-se pretender utilizar a frequência de 15 Hz que é claramente melhor, e continuar a experimentar novas configurações para pressão e velocidade com valores próximos a 30 psi e 40 m/s, respectivamente.

A experimentação contínua pode permitir aumentar ainda mais a produtividade da linha de montagem da lateral de porta.



14

Otimização de Respostas

Nível: Mestre

Otimização de Respostas

O que é otimização de respostas ?

Muitos experimentos envolvem a determinação de condições ótimas (configurações de fatores) que produzam o “melhor” valor para a resposta. Por exemplo, no desenvolvimento de um produto, pode ser necessário determinar as condições operacionais que resultem em um produto com propriedades desejáveis. Uma vez que cada propriedade é importante na determinação da qualidade do produto, é preciso considerar essas propriedades (respostas) simultaneamente. As configurações ótimas de fatores para uma resposta podem estar longe de serem ótimas para a outra resposta. A otimização simultânea de respostas é um método que permite chegar a um “acordo” entre as várias respostas.

Quando usar otimização de respostas ?

Utiliza-se a otimização simultânea de respostas para identificar a combinação de configurações de variáveis que otimize conjuntamente uma única resposta ou um conjunto de respostas. A otimização conjunta deve satisfazer aos requisitos para todas as respostas no conjunto.

Por que usar otimização simultânea de respostas ?

A otimização simultânea de respostas é utilizada para responder questões tais como :

- qual é o efeito de um conjunto de fatores sobre todas as respostas de interesse ?
- quais configurações para estes fatores asseguram que um processo satisfaça às especificações para todas as variáveis de resposta ?
- quais as configurações destes fatores otimizam simultaneamente todas as variáveis de resposta ?

Por exemplo :

- quais as configurações de tempo de reação, de temperatura da reação e do tipo de catalisador afetam o resultado e o custo da operação ?
- quais as configurações de tipo de material, de temperatura de injeção, de pressão de injeção e de temperatura de resfriamento maximizam o valor isolante, a densidade e a resistência do isolamento ?

Criando o experimento

Experimento

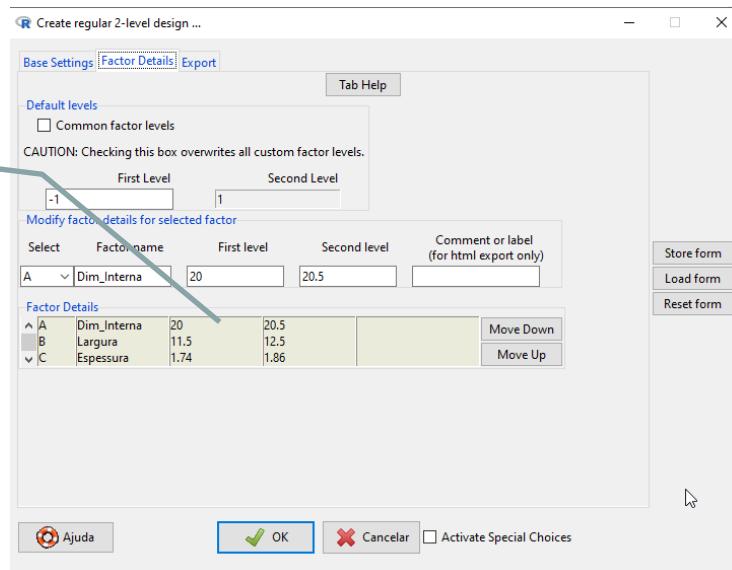
Quando o objetivo é otimizar simultaneamente duas ou mais respostas, utiliza-se o mesmo planejamento experimental para testar todas as respostas. Normalmente, as respostas são medidas simultaneamente.

Fatores

Os níveis do fator podem ser numéricos ou texto.

- os níveis numéricos são utilizados para variáveis quantitativas tais como : **Dimensão Interna, Largura e Espessura**.
- os níveis de texto são utilizados para variáveis categorizadas tais como por exemplo, **Tipo de máquina, adesivo A e B**, etc., de modo que o MINITAB saiba que o nível é qualitativo (neste exemplo, não existem fatores qualitativos).

Lembre-se,
use sempre
<ponto>
para
decimal.
<vírgula> é
outra coisa
em R.



Verificando o modelo

Veja abaixo: 16 corridas, 3 fatores, e que nenhum número está entre aspas, tipo: "11,2". Isso seria um texto...

```
RcmdrMsg: [2] NOTA: creating full factorial with 8 runs ...
Rcmdr> Design.1 <- FrF2(nruns= 16 ,nfactors= 3 ,blocks= 1 ,alias.block.2fis =
Rcmdr+   FALSE ,ncenter= 0 ,MaxC2 = FALSE ,resolution = NULL ,replications= 1 ,
Rcmdr+   repeat.only= FALSE ,randomize= FALSE ,seed= 13659 ,factor.names=list(
Rcmdr+     Dim_Interna=c(20,20.5),Largura=c(11.5,12.5),Espessura=c(1.74,1.86) ) )
Rcmdr> ## creator element of design.info will be different, when using the command line command!
RcmdrMsg: [3] NOTA: Os dados Design.1 tem 16 linhas e 4 colunas.
```

Espero que nessa altura do Campeonato não seja mais necessário colocar tela a tela do menu, certo? Avalie seu modelo se está tudo ok...

```
Rcmdr> print(Design.1)
      run.no run.no.std.rp Dim_Interna Largura Espessura Blocks
1       1           1.1        20    11.5    1.74    .1
2       2           2.1        20.5   11.5    1.74    .1
3       3           3.1        20    12.5    1.74    .1
4       4           4.1        20.5   12.5    1.74    .1
5       5           5.1        20    11.5    1.86    .1
6       6           6.1        20.5   11.5    1.86    .1
7       7           7.1        20    12.5    1.86    .1
8       8           8.1        20.5   12.5    1.86    .1
9       9           1.2        20    11.5    1.74    .2
10     10          2.2        20.5   11.5    1.74    .2
11     11          3.2        20    12.5    1.74    .2
12     12          4.2        20.5   12.5    1.74    .2
13     13          5.2        20    11.5    1.86    .2
14     14          6.2        20.5   11.5    1.86    .2
15     15          7.2        20    12.5    1.86    .2
16     16          8.2        20.5   12.5    1.86    .2
class=design, type= full factorial
NOTE: columns run.no and run.no.std.rp are annotation,
not part of the data frame
```

Interpretando os resultados

Gráfico de probabilidade normal

Se os efeitos dos fatores fossem zero e a variação devida apenas ao acaso, todos os pontos cairiam em uma linha quase reta.

Aqueles efeitos distantes da linha são considerados significativos, pois seus valores são de uma magnitude tal que a probabilidade deles ocorrerem por acaso é baixa.

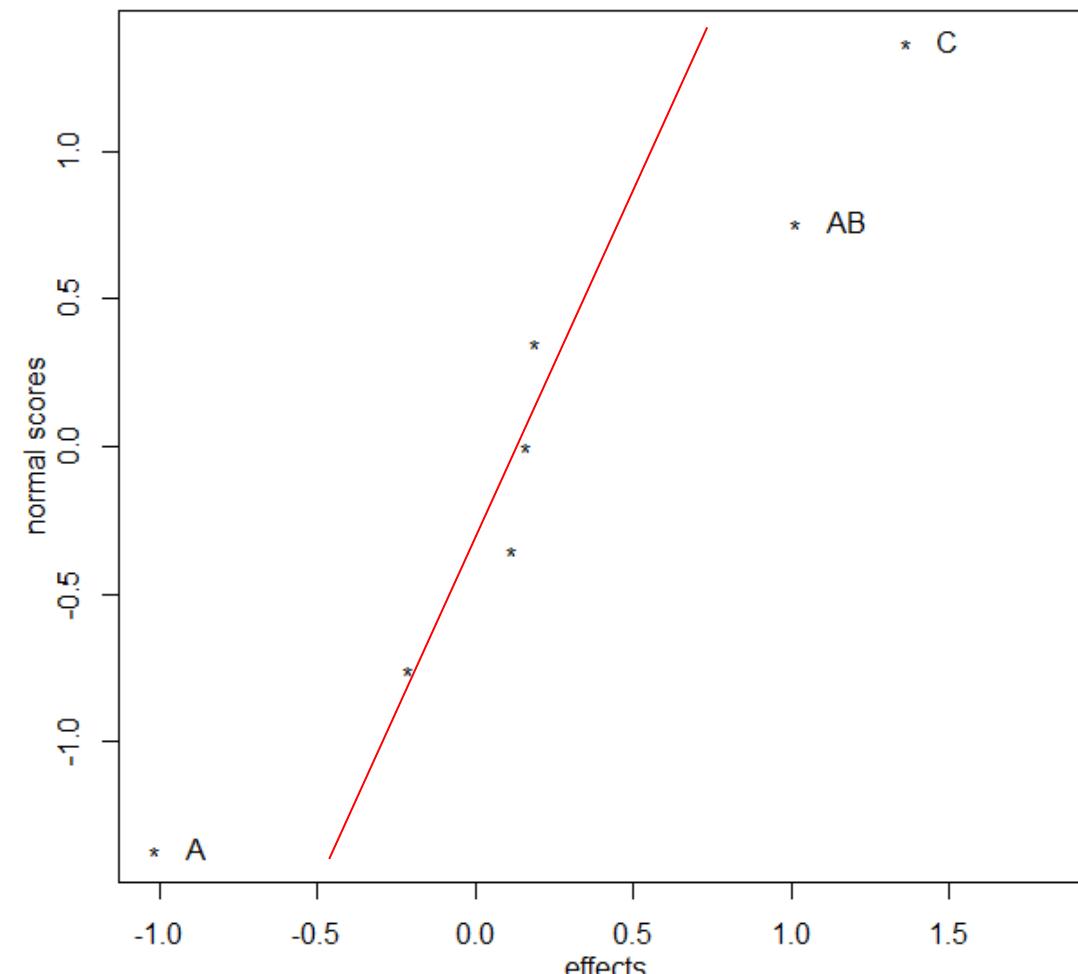
Aqui uma coisa interessante. O termo A é significativo. O termo B não, e o termo C sim.

Então podemos tirar o termo B, certo? Errado!. O termo B interage com o termo A, formando AB.

Logo nossa equação será:

$$A + C + A*B$$

Normal Plot for resposta3, alpha=0.1



A = Dim_Interna , B = Largura , C = Espessura

Interpretando os resultados

```

call:
lm.default(formula = resposta3 ~ Dim_Interna + Espessura + Dim_Interna * Largura, data = Design.1)

Residuals:
    Min      1Q Median      3Q     Max 
-1.3188 -0.1844  0.0625  0.1625  1.0813 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.04375   0.18650 10.958 0.000000294 ***
Dim_Internal -0.50625   0.18650 -2.714  0.0201 *  
Espessura1    0.68125   0.18650  3.653  0.0038 **  
Largura1      0.05625   0.18650  0.302  0.7686    
Dim_Internal:Largura1 0.50625   0.18650  2.714  0.0201 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.746 on 11 degrees of freedom
Multiple R-squared:  0.7192, Adjusted R-squared:  0.6171 
F-statistic: 7.042 on 4 and 11 DF,  p-value: 0.004587

Rcmdr> Anova(LinearModel.1, type="II")
Anova Table (Type II tests)

Response: resposta3
           Sum Sq Df F value    Pr(>F)    
Dim_Interna  4.1006  1 7.3681 0.020131 *  
Espessura    7.4256  1 13.3426 0.003802 ** 
Largura      0.0506  1 0.0910 0.768580    
Dim_Internal:Largura 4.1006  1 7.3681 0.020131 *  
Residuals    6.1219 11                        

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Interpretando os resultados

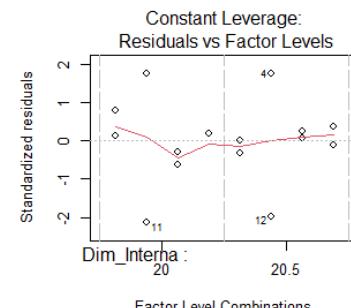
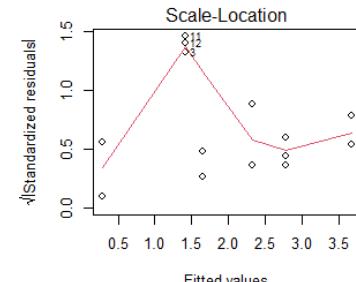
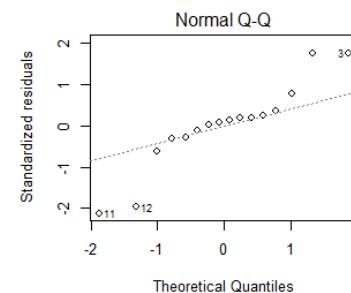
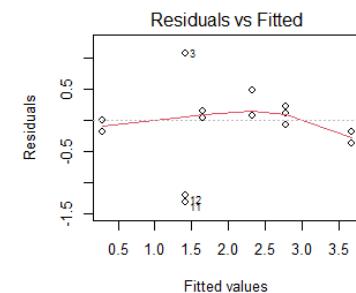
Ao lado Podemos perceber que o Rstudio retornou o valor da largura, mesmo que não tenhamos pedido, pois é um termo a ser usado na interação com Dim_Interna.

Temos os coeficientes para montar a equação de regressão.

O R quadrado ajustado é 61,7%

O valor-p da regressão é menor que 0,05

```
lm.default(resposta3 ~ Dim_Interna + Espessura + Dim_Interna * Largura)
```

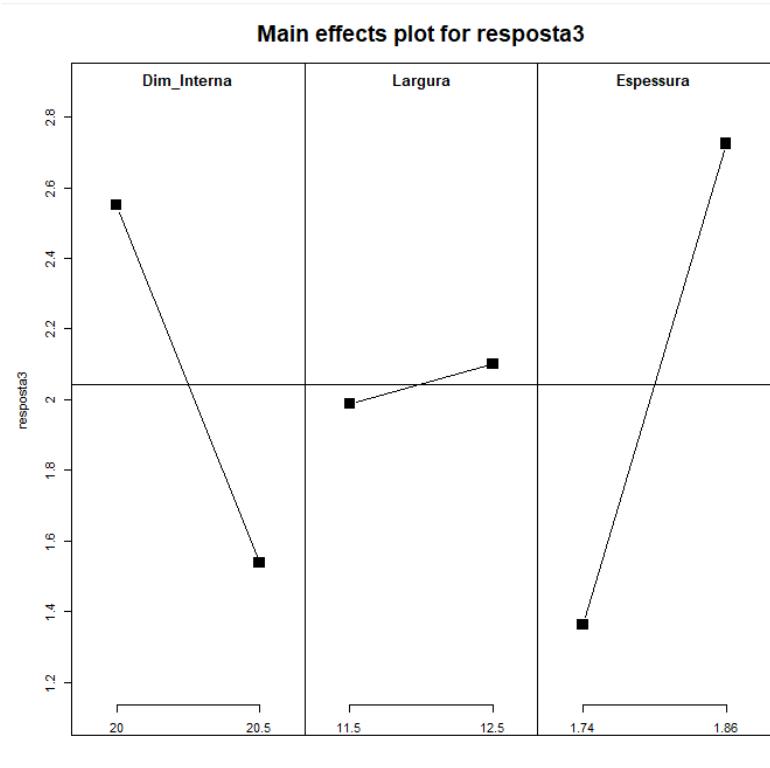


Interpretando os resultados

Gráfico dos efeitos principais

O gráfico de efeitos principais mostra que a melhor configuração para redução do desperdício de grampo C é :

- dimensão interna 20,5 (+1);
- largura 11,5 (-1), embora seja o item menos significativo;
- espessura 1,74 (-1).



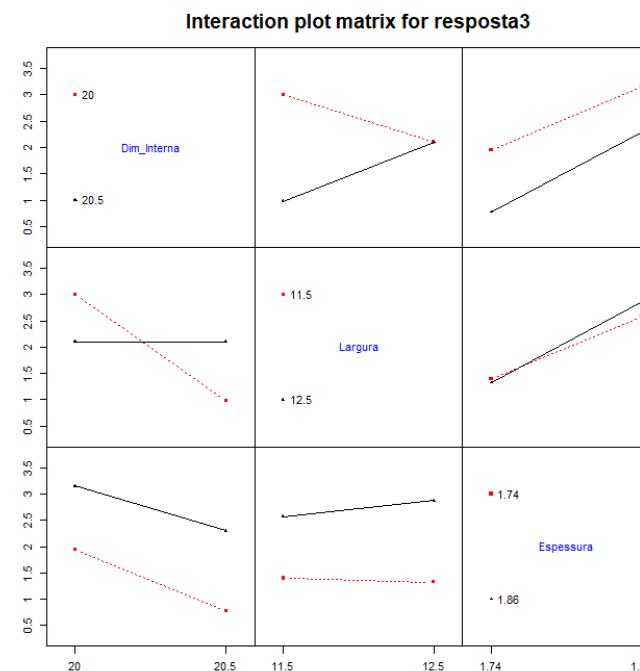
Interpretando os resultados

Gráfico de interações

O gráfico de interações mostra claramente que existe uma forte interação entre dimensão interna e largura, sendo que as melhores configurações são :

- interação dimensão interna (20,5 / +1) e largura (11,5 / -1);
- interação dimensão interna (20,5 / +1) e espessura (1,74 / +1);
- interação largura (11,5 / +1) e espessura (1,74 / -1).

Visualmente verificamos que não existe forte interação entre dimensão interna e espessura e entre largura e espessura, conforme já visto anteriormente através dos gráficos de probabilidade normal e Pareto.



Interpretando os resultados

Desejabilidade

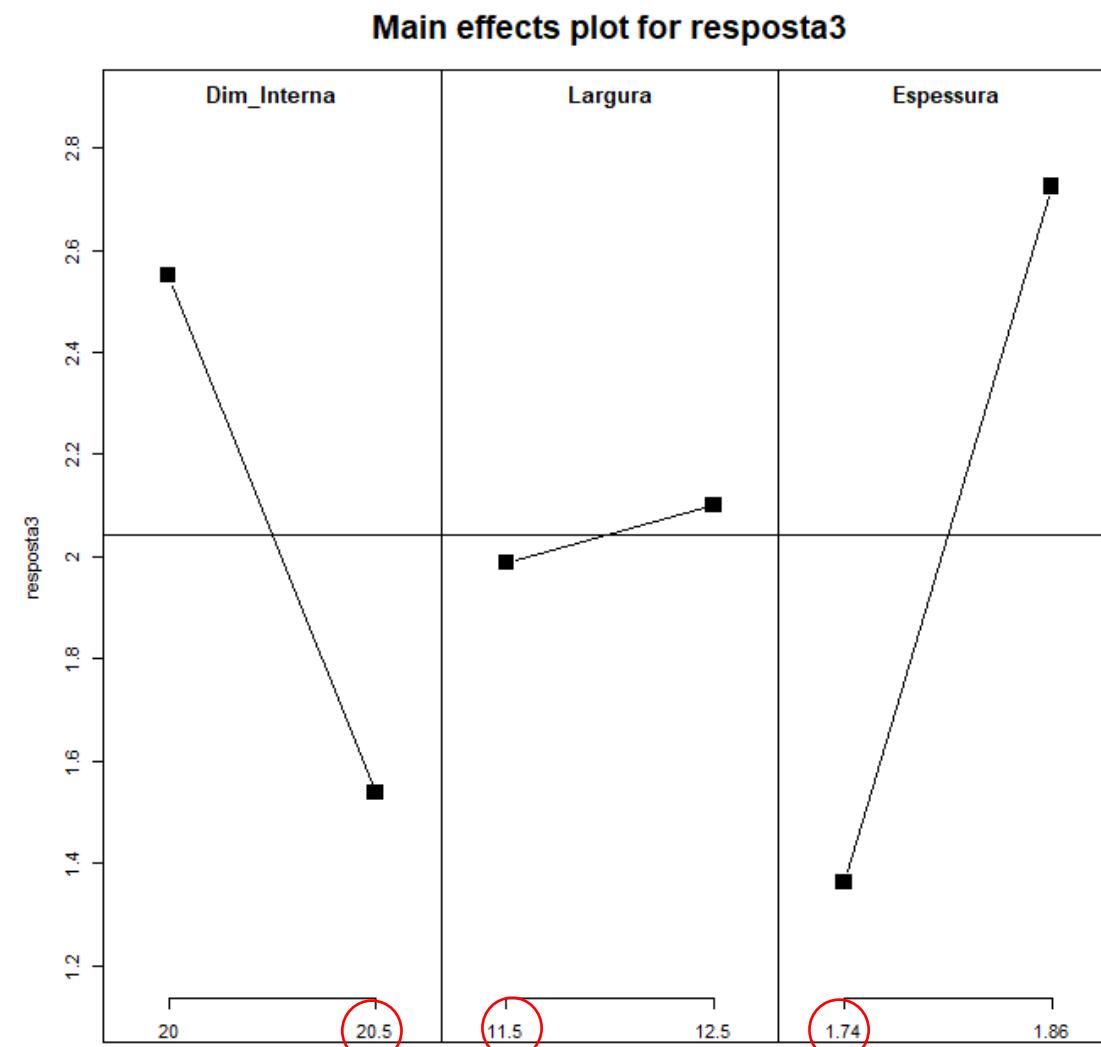
O MINITAB otimiza a desejabilidade geral ($D = 1,0000$). O MINITAB também exibe a desejabilidade individual de cada resposta :

- %Desperdício : $d = 1,0000$ (o valor é igual a D, pois só existe uma única resposta).

Os valores mínimos preditos da resposta ($y = 0,2938$) e as configurações necessárias para atingir estes resultados são os seguintes :

- Dim Interna = 20,5 (+1);
- Largura = 11,5 (-1);
- Espessura = 1,74 (-1).

Pode-se, também, observar o impacto de se mudar qualquer uma dessas configurações na resposta predita quando as linhas vermelhas do gráfico são movidas. Para retornar à condição otimizada, basta clicar em, localizado logo acima do gráfico.





15

Planejamento de Superfície de Resposta

Nível: Mestre

Planejamento de superfície de resposta

O que é um planejamento de superfície de resposta ?

Em um planejamento de superfície de resposta, os níveis dos fatores são variados para determinar qual condição de operação produz a resposta ótima ou desejada. É possível criar dois tipos de planejamentos de superfície de resposta no RStudio :

- planejamentos centrais compostos;
- planejamentos Box-Behnken.

Um planejamento de superfície de resposta é usado para modelar a relação entre uma variável resposta e dois ou mais fatores.

Para um modelo com dois fatores, a equação matemática que descreve esta relação é :

$$Y = \beta_0 + \beta_1 * A + \beta_2 * B + \beta_3 * A * B + \beta_4 * A^2 + \beta_5 * B^2 + \varepsilon,$$

onde A e B são os fatores de interesse.

A única diferença entre a equação de superfície de resposta e a equação para um planejamento fatorial é a adição dos termos quadrados (ou quadráticos). Esses termos são adicionados para modelar a curvatura na resposta.

Quando usar planejamento de superfície de resposta ?

Os planejamentos de superfície de resposta devem ser usados quando existir curvatura na variável de resposta. Isto geralmente ocorre quando as condições de operações dos fatores estão próximas a um valor ótimo da variável resposta.

Por que usar planejamento de superfície de resposta ?

Planejamentos de superfície de resposta podem ser usados para responder questões tais como :

- qual é o efeito de um conjunto de fatores em uma respostas de interesse ?
- quais condições de operação para estes fatores garantem um processo que satisfaça especificações desejadas ?
- quais condições de operação destes fatores otimizam a resposta ?

Por exemplo :

- qual é o efeito de açúcar e xarope de milho no sabor de um cereal ?
- quais são as condições de operação de temperatura e agente alvejante que resultarão em um papel com brilho dentro dos limites de especificação desejados ?
- quais são as condições de operação de tempo e temperatura que maximizam o resultado químico de um processo ?

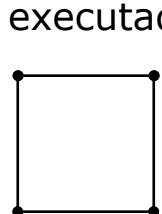
Planejamentos Centrais Compostos

O que é um planejamento central composto ?

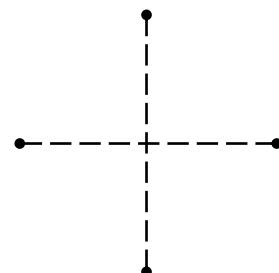
O planejamento central composto é o planejamento de superfície de resposta mais comumente usado. Os planejamentos centrais compostos consistem de :

- fatoriais 2^k ou 2^{k-1} ou pontos “cúbicos”, onde k é o número de fatores
- pontos axiais (também chamados pontos estrela)
- pontos centrais

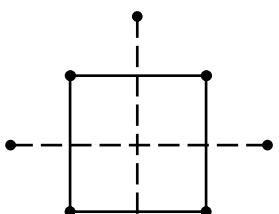
Abaixo, há um planejamento central composto com dois fatores. Os pontos nos diagramas representam os ensaios do experimento que foram executados :



Os pontos na porção factorial ou “cúbica” do planejamento são codificados como -1 e +1



Os pontos na porção axial ou estrela do planejamento, estão em $(+\alpha, 0)$, $(-\alpha, 0)$, $(0, +\alpha)$, $(0, -\alpha)$



Aparecem as porções “cúbica” (factorial) e axial com o ponto central. O centro do planejamento está em $(0,0)$

Quando usar um planejamento central composto ?

Use um planejamento central composto para :

- estimar de forma eficiente termos de primeira e segunda ordens
- modelar uma variável resposta com uma curvatura acrescentando pontos a um planejamento fatorial executado anteriormente

Por que usar um planejamento central composto ?

Um planejamento central composto pode ser usado para responder a questões tais como :

- qual é o efeito de um conjunto de fatores em uma resposta ?
- quais condições de operação para esses fatores asseguram que que satisfaça as especificações desejadas ?
- quais condições de operação destes fatores otimizam a resposta ?

Por exemplo :

- qual é o efeito de açúcar e xarope de milho no sabor de um cereal ?
- quais são as condições de operação de temperatura e agente alvejante que resultarão em um papel com brilho dentro dos limites de especificação desejados ?
- quais condições de operação de tempo e temperatura maximizam o resultado químico de um processo ?

Criando um planejamento central composto

O objetivo desta análise é identificar a combinação de frequência (Hz) e pressão (psi) que otimize (minimize) a geração de defeitos na linha Door Panel.

Variáveis

Inicie a análise criando um planejamento central composto para avaliar o impacto da frequência e da pressão na geração de defeitos na linha Door Panel.

Planejamentos

Para um planejamento central composto com dois fatores, pode-se escolher entre :

- um planejamento de 13 ensaios com um bloco
- um planejamento de 14 ensaios com dois blocos

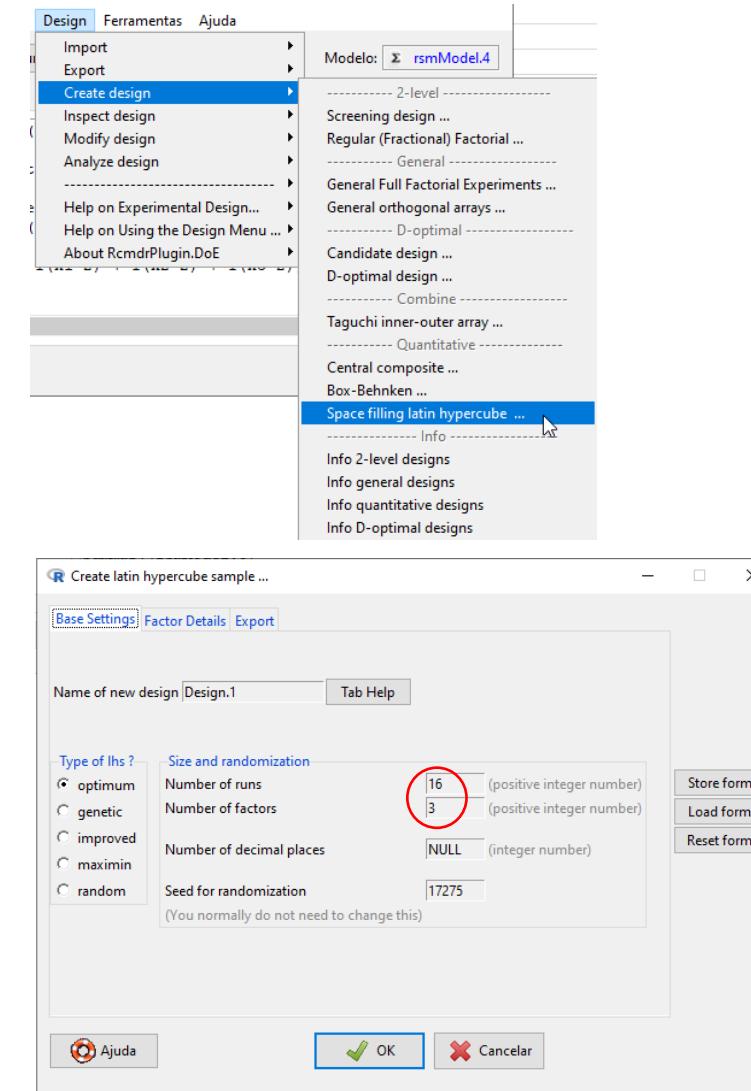
Valor de alfa

O Rstudio determina valores a default para assegurar que o planejamento exiba as propriedades desejadas de blocos ortogonais e, sempre que possível, rotatibilidade.

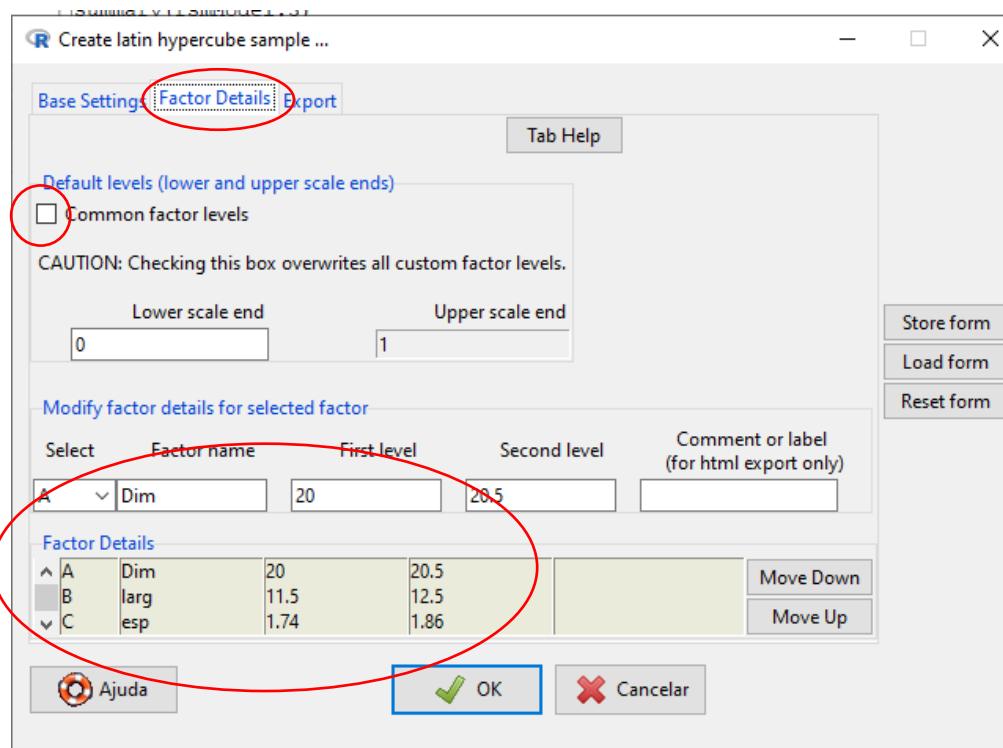
Com um planejamento em blocos ortogonais, os termos lineares e de interação no modelo podem ser estimados independentes uns dos outros.

Com um planejamento rotacionável, todos os pontos equidistantes do centro do experimento tem a mesma variância de predição.

Criar um planejamento de superfície de resposta



Planejamento de Superfície de Resposta



```
Rcmdr> Design.1 <- lhs.design( type= "optimum" , nruns= 16 ,nfactors= 3 ,digits= NULL ,seed= 17849 ,  
Rcmdr+   factor.names=list( Dim=c(20,20.5),larg=c(11.5,12.5),esp=c(1.74,1.86) ) )
```

Veja no commander o nome do modelo e substitua no comando.
Aqui é rsmModel.4

Interpretação dos resultados

```
Rcmdr> print( Design.1 )  
Dim      larg      esp  
1 20.30483 12.46462 1.833357  
2 20.28003 11.79726 1.829433  
3 20.02263 12.36185 1.815163  
4 20.07279 11.74281 1.839251  
5 20.42932 11.84824 1.846135  
6 20.49483 11.66478 1.812230  
7 20.16357 11.50425 1.785130  
8 20.35713 11.61172 1.772916  
9 20.22094 12.29432 1.800074  
10 20.15287 11.88619 1.745120  
11 20.10093 12.24524 1.763266  
12 20.05197 11.95210 1.783907  
13 20.32050 12.13056 1.755734  
14 20.21107 12.03764 1.854115  
15 20.38290 12.42756 1.754436  
16 20.45787 12.08756 1.797876  
class=design, type= lhs
```

Resumo do Planejamento

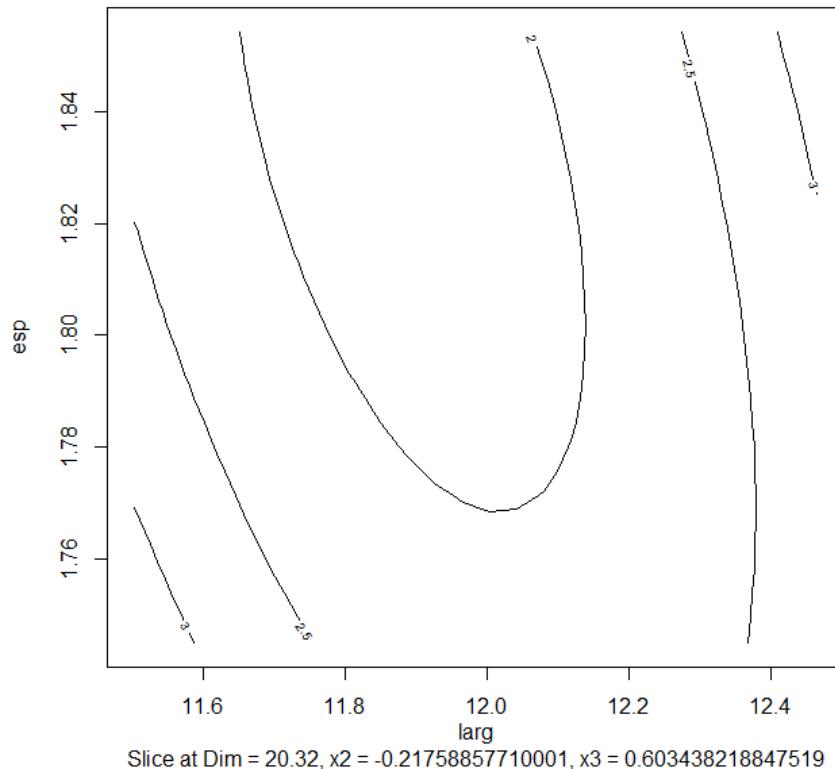
Para o planejamento com o cereal, foi criado um planejamento com os valores acima. Agora vamos analisar

```
R Commander  
Arquivo Editar Dados Estatísticas Gráficos Modelos Distribuições Design Ferramentas Ajuda  
Conjunto de Dados: Design.1.coded.coded Editar conjunto de dados Ver conjunto de dados Modelo: rsmModel.4  
R Script R Markdown
```

```
234 #Superfície de resposta  
235 install.packages("srn")  
236 library("srn")  
237 contour(rsmModel.4, ~x1+x2+x3, at = xs(rsmModel.4))  
238 persp(rsmModel.4, ~x1+x2+x3, at = xs(rsmModel.4))  
239
```

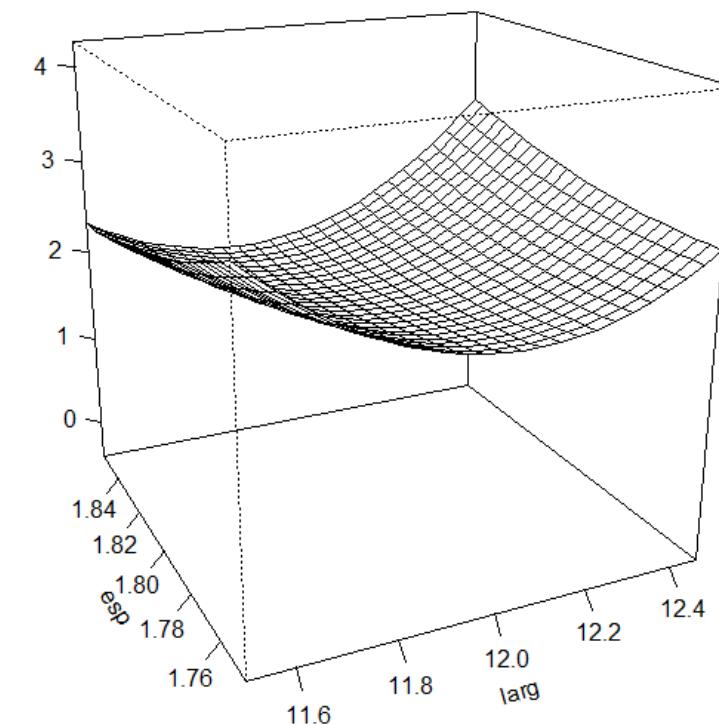
Encontrando as condições ótimas

O objetivo desta análise é identificar a combinação da frequência e da pressão que otimize a quantidade de defeitos na linha Door Panel. É necessário localizar a região ou regiões dos níveis dos fatores que otimize este número. O gráfico de contorno mostra algumas regiões ótimas (entre 2 e 3 defeitos).



Encontrando as condições ótimas

O gráfico de superfície (3D) mostra a mesma informação que o gráfico de contorno, porém com outro formato.



Considerações finais

Conclusões práticas

As condições ótimas para frequência e pressão encontram-se nos extremos :

- o número de defeitos foi minimizado quando a frequência e a pressão estão em suas configurações mínimas;
- outra ponto ótimo ocorre quando frequência e pressão estão em suas configurações máximas;

Além de investigar os efeitos da frequência do gerador de ultra-som e da pressão do ar comprimido na geração de defeitos na linha Door Panel, pode-se estar interessado em :

- executar um planejamento de superfície de resposta que não configure os fatores nos níveis baixos simultaneamente;
- executar um planejamento de misturas para entender melhor a dependência entre frequência e pressão.

Considerações estatísticas

Com um planejamento central composto, deve-se executar combinações extremas dos níveis mesmo quando estes níveis não fazem sentido em termos práticos. Por exemplo, a minimização dos defeitos na linha Door Panel ocorreu quando :

- frequência e pressão estavam em suas configurações mais baixas;
- frequência e pressão estavam em suas configurações mais altas.

Poderia ter escolhido um planejamento que não incluisse estas condições por que há poucas probabilidades que elas otimizem a quantidade de defeitos na linha Door Panel (lembrem-se da baixa taxa de explicação da variação existente).

Os planejamentos centrais compostos englobam os ensaios nos níveis dos fatores no mínimo e máximo especificado.

Existem duas soluções possíveis para este problema :

- especificar as condições dos fatores mínima e máxima para os pontos axiais e deixe o RStudio ajustar os pontos cúbicos;
- usar um planejamento central composto face-centered.

Estes planejamentos não forçam os pontos axiais além das condições mínima e máxima especificada para cada fator.



16

Cartas de Controle (Variáveis)

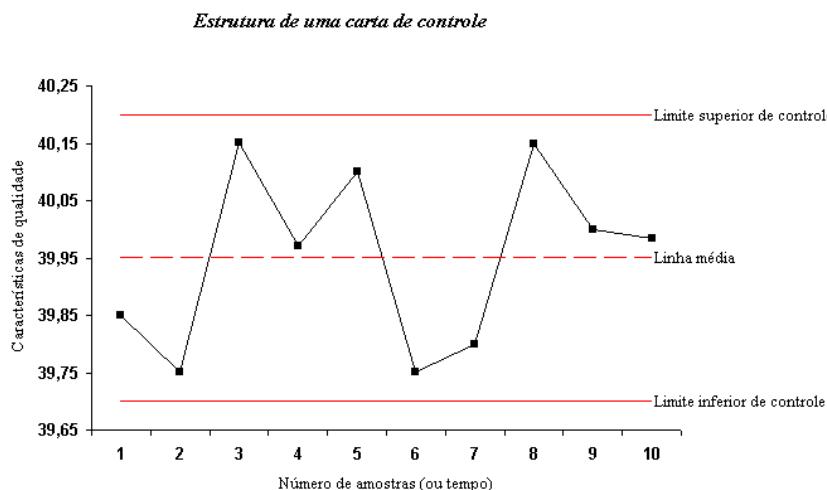
Nível: Avançado

Cartas de controle

O que é uma carta de controle ?

Uma carta de controle é uma carta especial de uma série temporal cujo objetivo é identificar padrões não normais de variabilidade no processo. As cartas de controle X e R são, provavelmente, as cartas de controle mais comumente usadas.

Abaixo pode-se ver uma carta de controle genérica.



Quando usar uma carta de controle ?

Utilize uma carta de controle quando tiver dados de amostras ordenados por tempo para uma característica de qualidade importante.

Por que usar uma carta de controle ?

Cartas de controle podem ajudá-lo a responder questões como, por exemplo :

- Quão consistente é o produto ?
- Quanta variação há para cada operador ou dentro de cada lote ?
- Quanta variabilidade há entre operadores ou lotes ?

Por exemplo :

- Os novos operadores estão recebendo suficiente treinamento, de forma que seu trabalho seja consistente ?
- As máquinas estão ajustadas de forma similar, de modo que seus produtos sejam similares ?

Cartas de Controle (Variáveis)

Carta X-bar - R

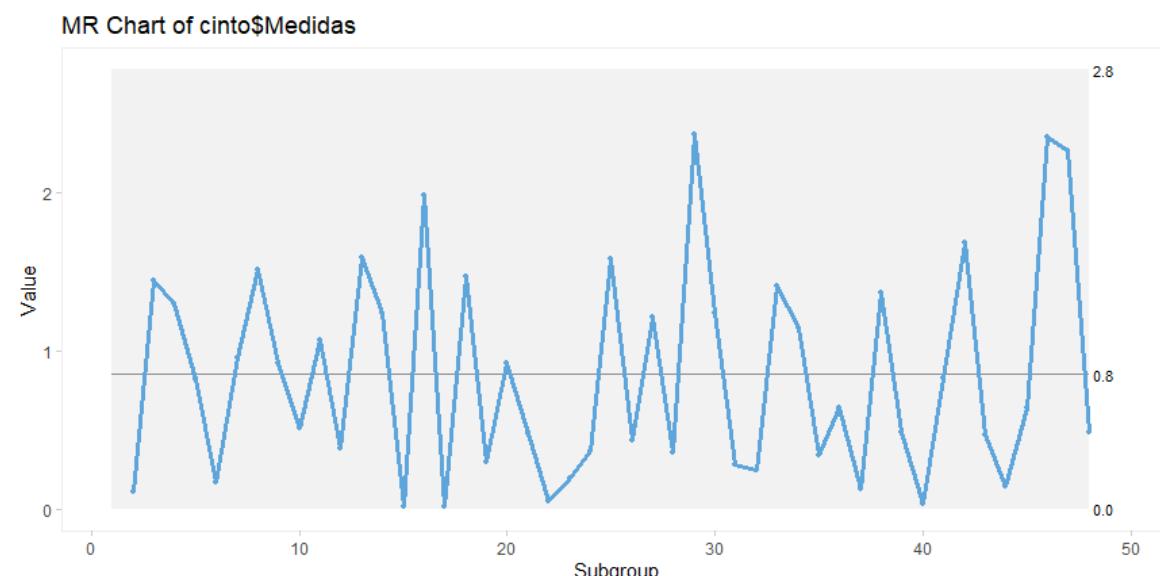
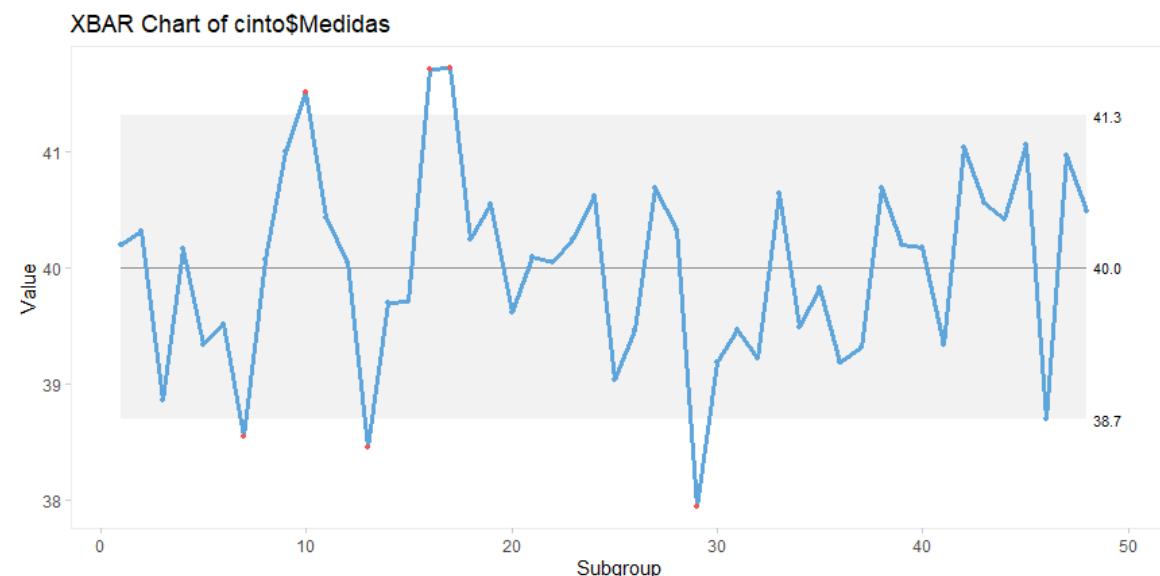
Seguem as definições :

Xbar : Uma carta X permite determinar se a média do processo (centro ou localização) está sob controle.

R : Uma carta R é uma carta da amplitude dos valores dentro de cada subgrupo de dados ordenados no tempo

Com a carta Xbar / R, é possível unificar estas definições e consequentemente os estudos a fim de facilitar nosso trabalho. A primeira análise deve ser feita à carta de amplitude (R), e esta estando OK, a análise se extenderá para a carta de médias (X-bar).

```
240 #cartas de Controle
241 # x-barra e R
242 install.packages("qicharts2")
243 library("qicharts2")
244 cinto<- read_excel("CINTOSEG.xlsx")
245 help("qic")
246 view(cinto)
247 qic(cinto$Medidas, chart = "xbar", x = cinto$Grupo)
248 qic(cinto$Medidas, chart = "mr", x = cinto$Grupo)
```



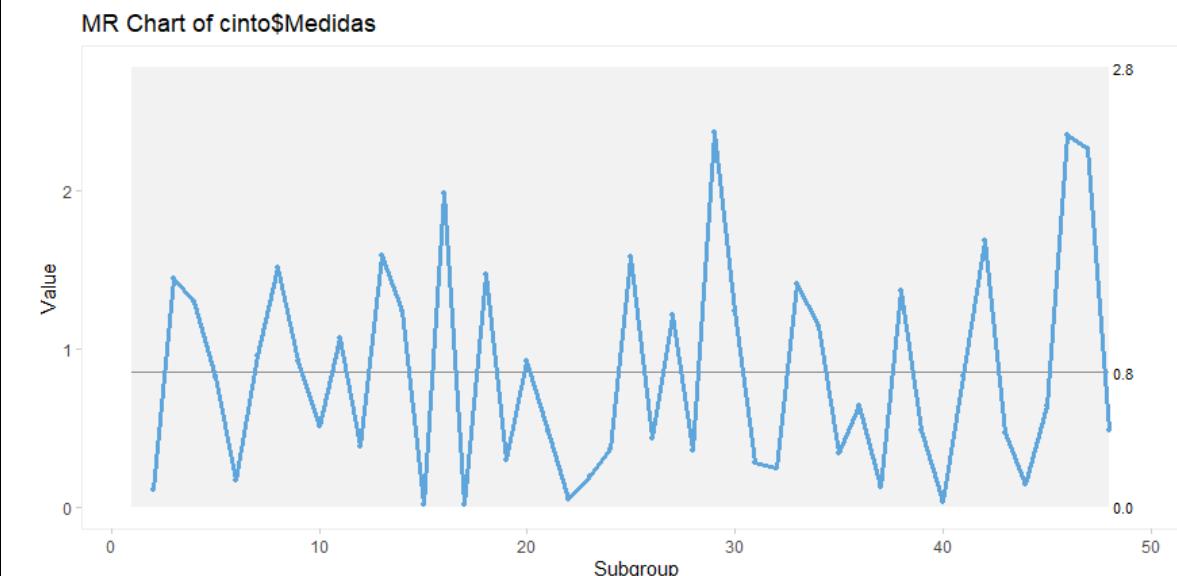
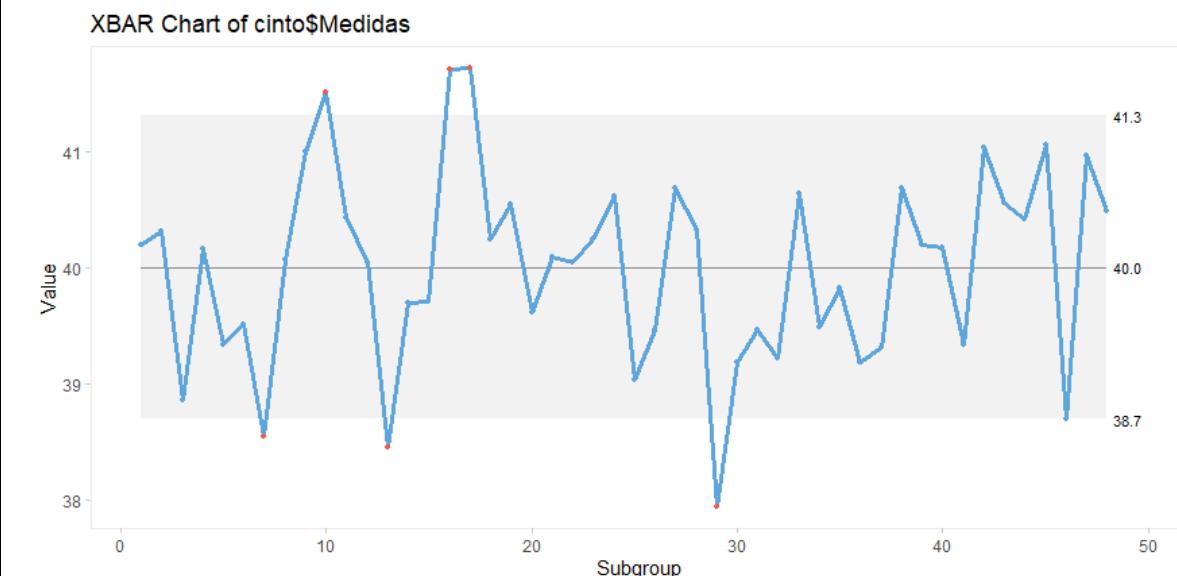
Interpretando os resultados

Limites de controle

Limites de controle são calculados usando todos os dados da amostra. Isto cria um problema quando você está trabalhando em melhoria de processos, por que a mudança de média e variação irão fazer com que muitos pontos fiquem fora de controle.

Gerando uma carta de melhoria de processo

Quando houver mudanças no processo, é interessante que se recalcule os limites de controle.



Carta Xbar / S

O que é uma carta S ?

Uma carta S é uma carta de controle para os desvios padrão de subgrupos.

Quando usar uma carta S ?

Utilize uma carta S quando desejar comparar a variabilidade que está ocorrendo dentro de cada subgrupo amostral.

Sempre se deve checar uma carta R ou uma carta S antes de tentar interpretar uma carta das médias dos subgrupos (uma carta X). Se a variabilidade nos subgrupos não for estável, os limites de controle usados no gráfico X não serão significativos.

Cartas S são tipicamente usados para investigar a variação de processo em amostras maiores do que 10, enquanto cartas R são usados para amostras menores.



Por que usar uma carta S ?

Uma carta S pode ser usada para responder questões tais como :

- Os operadores estão desempenhando suas funções de forma consistente ?
- As máquinas estão desempenhando suas funções de forma consistente ?

Por exemplo :

- Os operadores de furadeiras de coluna estão fazendo perfurações de forma consistente ?
- Os tornos mecânicos estão produzindo peças com diâmetros consistentes ?

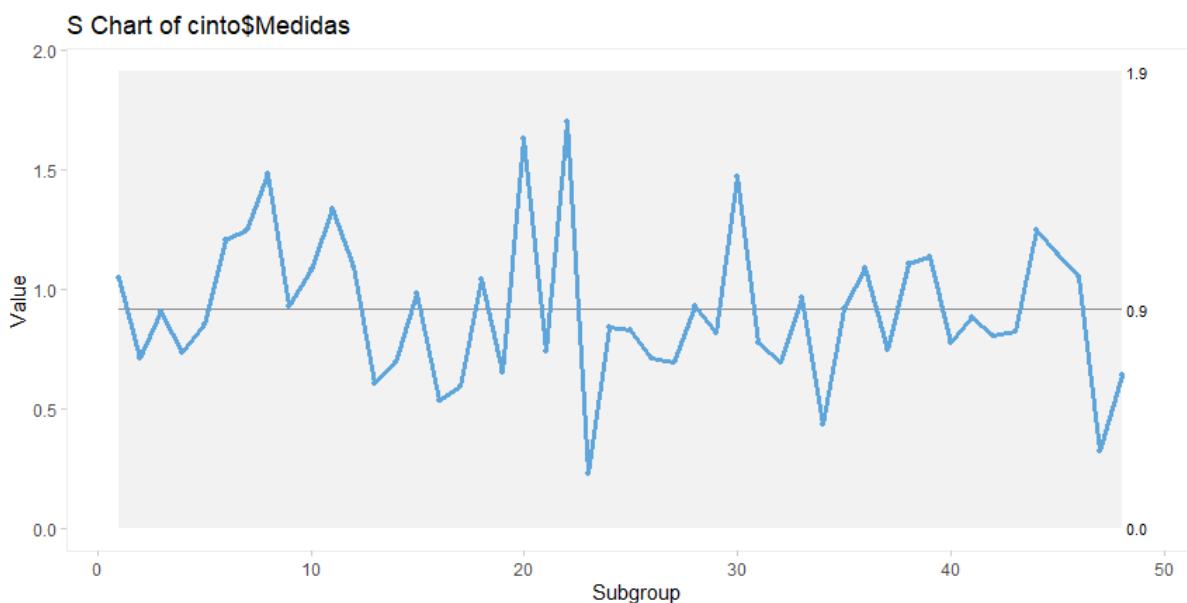
Criando uma carta S

Escolhendo uma carta para variação por cada subgrupo

Para este exemplo, uma carta Xbar-S e não Xbar-R será usada.

Isto porque o tamanho do subgrupo é maior que 10.

```
249 qic(cinto$Medidas, chart = "s", x = cinto$Grupo)
```



Interpretando os resultados

Limites de controle

Limites de controle são calculados usando todos os dados da amostra. Isto cria um problema quando você está trabalhando em melhoria de processos, por que a mudança de média e variação irão fazer com que muitos pontos fiquem fora de controle.

Gerando uma carta de melhoria de processo

Quando houver mudanças no processo, é interessante que se recalcule os limites de controle.

Considerações finais

Conclusões práticas

Parece que é possível reduzir a variação no processo de torqueamento do parafuso de fixação do suporte do cinto de segurança.

Para visualizar o processo antes e depois da mudança, recalcule os limites de controle assim que as alterações forem implantadas.

Considerações estatísticas

Utilize uma carta X e S para monitorar e detectar mudanças na variação do processo e na média dos dados quando estes estiverem em subgrupos maiores do que 10. Para subgrupos de menor tamanho, utilize cartas X e R.

Quando houver uma mudança no processo e for necessário recalcular os limites de controle, pode-se criar cartas de controle históricas para fazer o acompanhamento da melhoria do processo. É importante recalcular os limites de controle quando é feita uma mudança no processo.

Cartas de Controle (Variáveis)

Carta X-AM (I-MR)

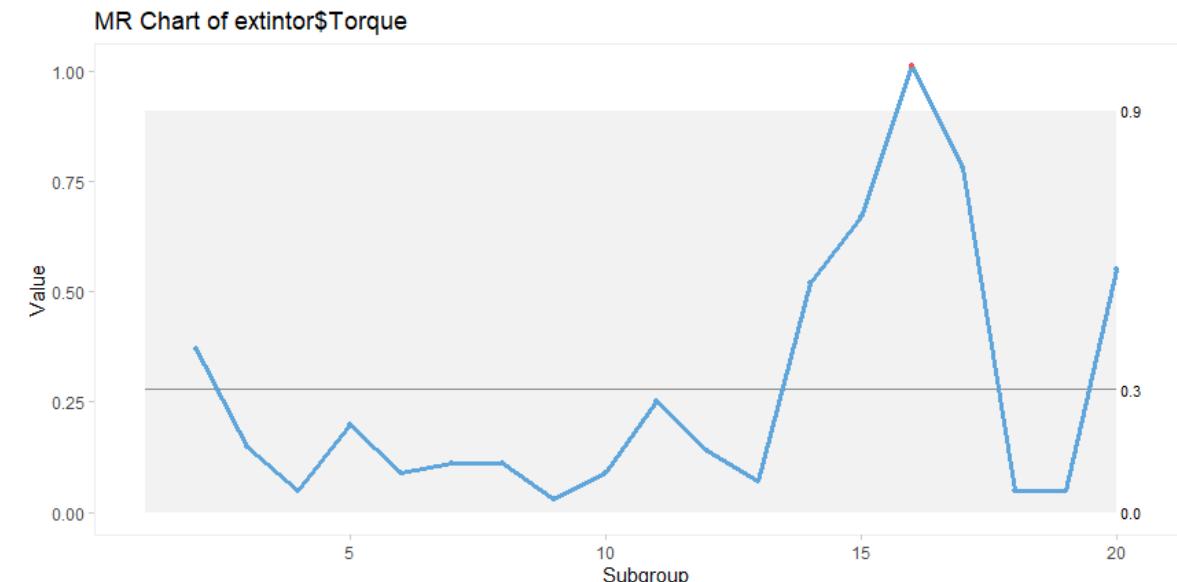
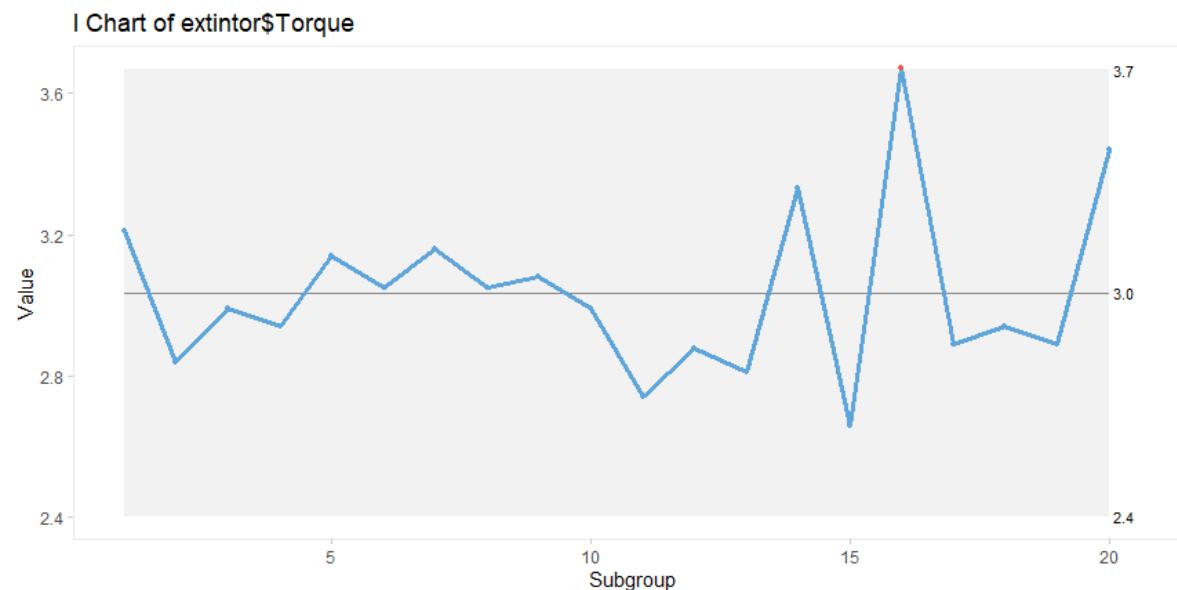
Seguem as definições :

X (I): A carta de valores individuais (X) permite determinar se a média do processo (centro ou localização) está sob controle quando os dados são medidas individuais ao invés de subgrupos.

AM (MR) : A carta de amplitudes móveis (AM) permite monitorar e detectar mudanças na variação de um processo quando os dados são medidas individuais.

Com a carta X / AM (I / MR), é possível unificar estas definições e consequentemente os estudos a fim de facilitar nosso trabalho.

```
251 # I-MR
252 extintor<- read_excel("EXTINTOR.xlsx")
253 qic(extintor$Torque, chart = "i")
254 qic(extintor$Torque, chart = "mr")
```



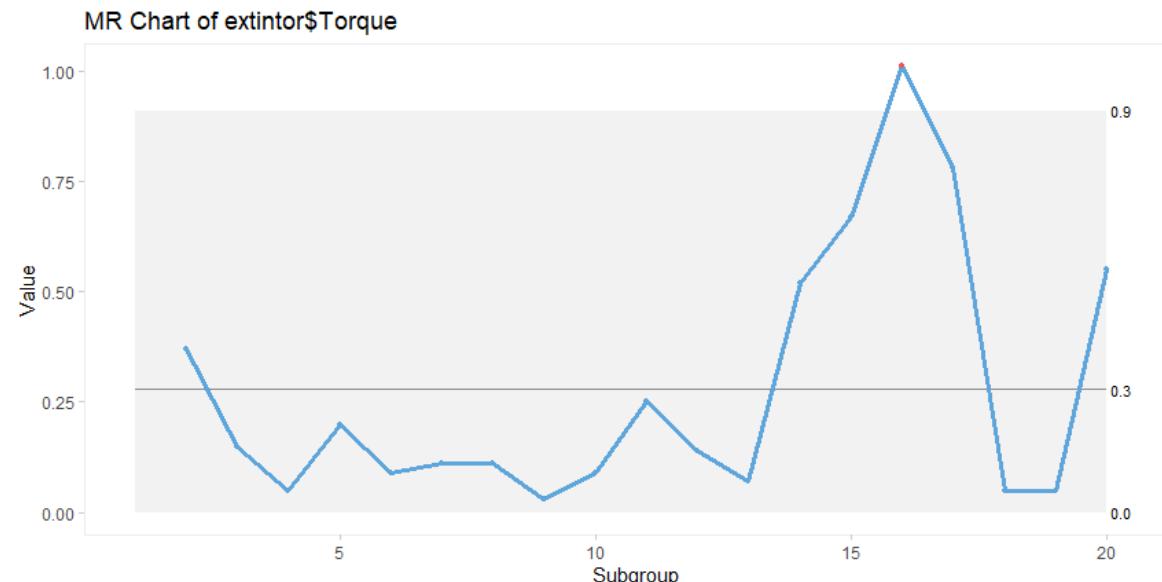
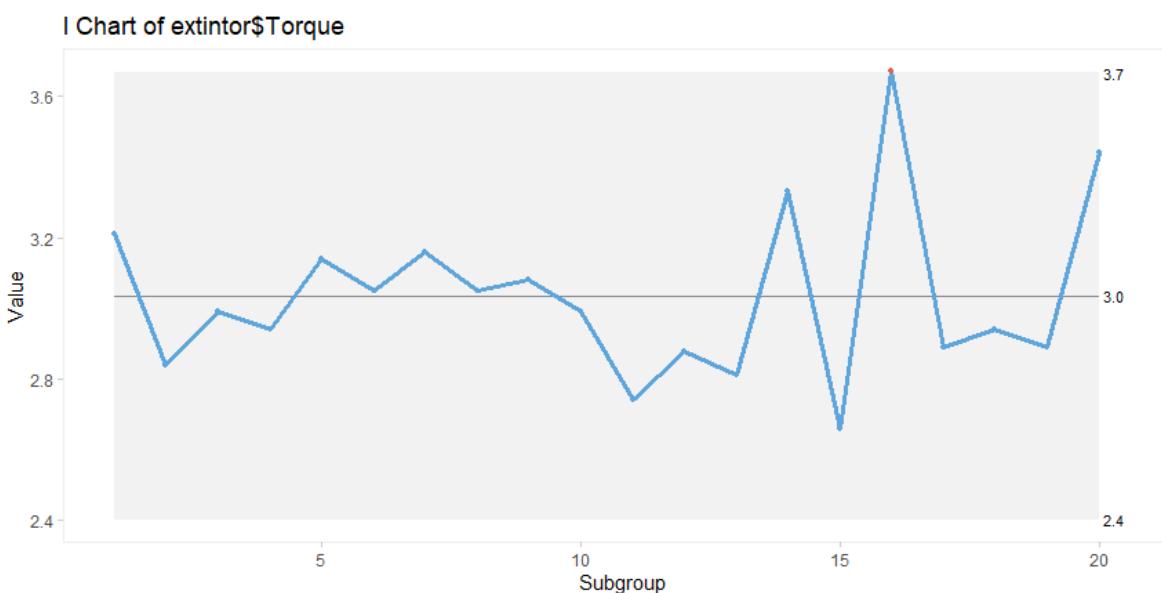
Interpretando os resultados

Limites de controle

Limites de controle são calculados usando todos os dados da amostra. Isto cria um problema quando você está trabalhando em melhoria de processos, por que a mudança de média e variação irão fazer com que muitos pontos fiquem fora de controle.

Gerando uma carta de melhoria de processo

Quando houver mudanças no processo, é interessante que se recalcule os limites de controle.





17

Cartas de Controle (Atributos)

Nível: Avançado

Atributos

O que são atributos ?

Atributos descrevem características de qualidade que só podem ser definidas pela presença ou ausência de um dado defeito. Exemplos de atributos incluem rachaduras em uma peça de cerâmica, descamações em uma superfície pintada, empenamento em material de construção e erros tipográficos em um documento. Os termos a seguir são usados para descrever dados desse tipo :

- **Unidade defeituosa** : quando o produto deixa de atender a requisitos devido à presença de uma não conformidade, a unidade é declarada inutilizável e considerada como defeituosa. Os dados só podem assumir um dentre dois valores, tais como aprovado/reprovado, vai/não vai ou presente/ausente. Por exemplo : se a rosca em um parafuso não é aprovada em uma inspeção, a unidade inteira é descartada. Dados desse tipo geralmente seguem uma distribuição **Binomial**;
- **Não conformidade** (também chamada de defeito) : a não conformidade descreve um defeito ou mais em uma única unidade. Por exemplo : uma vestimenta pode apresentar descoloração, ter botões faltando, ou uma costura imperfeita. Por existir mais de uma não conformidade em uma unidade, os dados coletados são contados e, geralmente, seguem uma distribuição de **Poisson**.

Quando usar atributos

Atributos são úteis :

- Quando é difícil representar numericamente as características de qualidade. Nessas situações, os itens podem ser classificados como conforme ou não conforme às especificações. Por exemplo : uma placa de circuitos tem um componente faltando ou não e uma lâmpada se acende ou não;
- Quando se deseja considerar várias características de qualidade em conjunto. Se cada característica for considerada uma variável, é preciso analisá-las em separado. Por exemplo, se considerarmos arranhões, descamações e bolhas em uma superfície pintada como atributos, podemos analisá-las em conjunto e tratá-las como defeitos;
- Quando os dados estão prontamente disponíveis : inspeções, registros de serviço, solicitações de reparos ou relatórios administrativos, já feitos.

Por que usar atributos ?

Usando atributos, você pode responder aos seguintes tipos de questões :

- O número de unidades arranhadas é estável ao longo da produção ?
- O número de não conformidades diminui à medida que o processo funciona por mais tempo ?

Carta P

O que é uma carta P ?

A carta P monitora a proporção de unidades defeituosas no processo. Para criar esse gráfico, cada unidade é inspecionada e declarada conforme ou não conforme. A proporção de unidades defeituosas (P) em cada subgrupo aparece como o ponto plotado na carta.

Para cada subgrupo, P é calculado por :

$$p = \frac{\text{nº unidades não conformes}}{\text{tamanho do subgrupo}}$$

Quando usar uma carta P ?

Utilize a carta P quando o item de controle for a proporção de unidades defeituosas.

Por que usar uma carta P ?

Utilizar uma carta P para avaliar se o processo está sob controle. A carta P pode ser usada para responder questões tais como :

- erros em transações bancárias são estáveis com o decorrer dos dias, quando o número de transações muda ?
- o número de lâmpadas defeituosas varia quando o número da amostra altera ?
- a proporção de clientes insatisfeitos muda com as estações ?

Cartas de Controle (Atributos)

Avaliando a proporção de bancos não OK

O objetivo é minimizar a quantidade de bancos que são retrabalhados na célula de retrabalho diariamente. Utilize a carta P para gerar um gráfico da proporção de unidades defeituosas nos subgrupos.

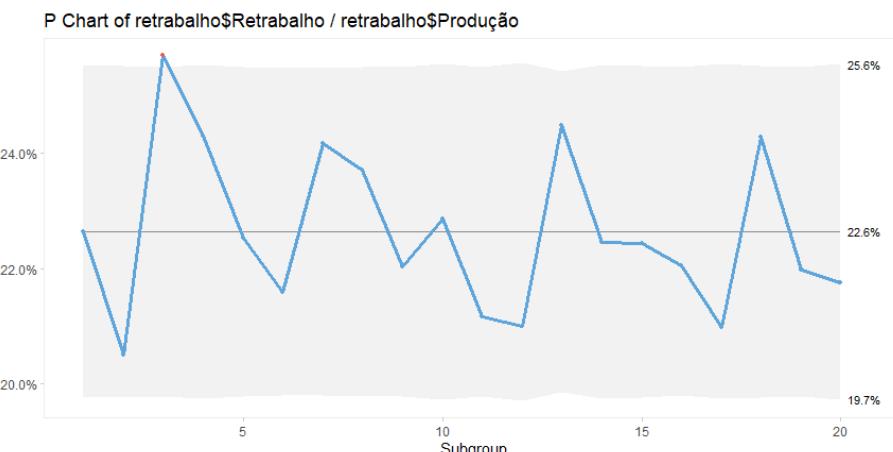
Variáveis

As respostas estão dispostas em três colunas na planilha; aqui serão utilizadas as seguintes :

- **Data** contém as datas em que o controle foi realizado
- **Produção** contém o volume diário de bancos produzidos
- **Retrabalho** contém o volume diário de bancos retrabalhados

Testes

Quando os tamanhos dos subgrupos são desiguais, os limites de controle não são constantes.



Carta P

```
256 #p
257 retrabalho<- read_excel("RETRABALHO.xlsx")
258 qic(retrabalho$Retrabalho, chart = "p", n=retrabalho$Produção)
259
```

Interpretação dos resultados

A carta P mostra o número de amostras no eixo x e a proporção de unidades defeituosas no eixo y. Os pontos plotados representam a proporção de unidades defeituosas em cada subgrupo. P é uma estimativa do percentual geral de unidades defeituosas.

A proporção média de bancos retrabalhados diariamente é de 0,22643, Multiplica-se a proporção por 100 para obter o percentual de bancos retrabalhados diariamente.

$$0,22643 \times 100 = 22,643\%$$

Em média, aproximadamente 23% dos bancos produzidos são enviados para o posto de retrabalho diariamente. Isso é um volume inaceitável de retrabalho de bancos.

Limites de controle

Uma vez que os limites de controle são em função do tamanho do subgrupo, eles flutuam sempre que o tamanho do subgrupo muda. Quando o tamanho do subgrupo aumenta, os limites de controle se movem para mais perto da linha central porque o percentual de unidades defeituosas em um subgrupo (P) estará mais próximo da média do processo (P).

Considerações finais

Conclusões práticas

Em média, 23% dos bancos produzidos passam pelo posto de retrabalho diariamente. Isso gera um potencial muito grande de enviarmos material não conforme para nossos clientes.

O próximo passo poderia ser determinar as razões para o retrabalho de bancos e definir uma meta para este tipo de ocorrência.

Retrabalho Zero é possível ?

Se for considerado importante chegar a um índice de 0% de bancos retrabalhados, pense nos custos envolvidos.

Considerações estatísticas

Utilize uma carta P para avaliar se um processo está sob controle, quando o item de controle for a proporção de unidades defeituosas.

Lembre-se, “sob controle” não quer dizer que todas as unidades estão sem defeitos, mas apenas que o processo está produzindo saída de uma forma consistente.

Embora atributos possam ser mais convenientes de coletar do que variáveis, considere o seguinte :

- atributos indicam apenas se uma peça ou serviço atende às especificações; eles não indicam a extensão em que as especificações estão sendo atendidas ou não;
- atributos só são capazes de indicar se o índice de não conformidade está mudando. Variáveis contém informações sobre o centro e a dispersão da distribuição e podem indicar uma mudança no processo antes que o índice de não conformidades aumente;
- cartas de controle para atributos requerem amostras de tamanho maior do que as cartas de controle para variáveis;
- quando os subgrupos são muito grandes (500 ou mais), os limites de controle ficam muito próximos da linha central. Neste caso, a maioria dos pontos parecerá estar fora de controle.

Carta NP

O que é uma carta NP ?

A carta NP monitora o número de unidades defeituosas em um processo. Para criar esta carta, cada unidade é inspecionada e declarada conforme ou não conforme. O número de unidades defeituosas em cada subgrupo é um ponto na carta NP.

Quando usar uma carta NP ?

Utilize a carta NP quando o item de controle for o número de unidades defeituosas e o tamanho do grupo for constante.

Se os tamanhos dos subgrupos variar, utilize a carta P, a qual mapeia a proporção de unidades defeituosas.

Por que usar uma carta NP ?

Utilize uma carta NP para avaliar se o processo está sob controle. A carta NP pode ser usada para responder questões tais como :

- o número de microchips defeituosos é estável ?
- o processo de fabricação de chapas metálicas de rodas de bicicletas está sob controle ?
- o nº de painéis de compensado muda ao longo da semana ?

Cartas de Controle (Atributos)

Avaliando o índice de painéis não OK

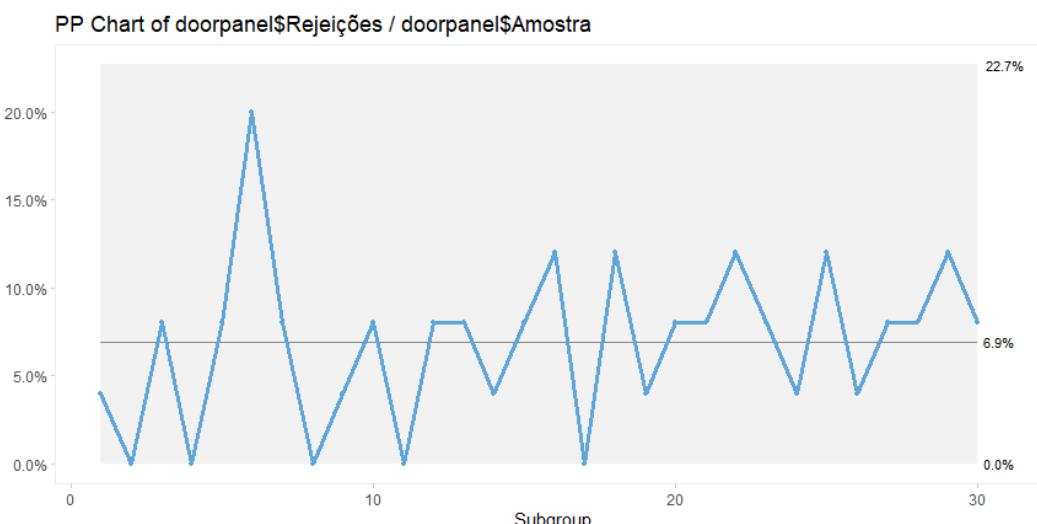
Para ajudar a melhorar a qualidade dos nossos painéis de porta, estamos implementando um novo sistema de inspeção. Utilize a carta NP para avaliar o número de unidades defeituosas (rejeições).

Variáveis

As respostas são dispostas em uma única coluna (**Rejeições**), a qual contém o número de unidades defeituosas para cada subgrupo. Uma vez que o tamanho do subgrupo é constante, não é preciso uma coluna para indicar o tamanho do subgrupo.

Testes

Quando o tamanho dos subgrupos é o mesmo (constante), os limites de controle são constantes. Aqui também é possível a realização de testes à procura de causas especiais.



Carta NP

```
260 #np
261 doorpanel<- read_excel("DOORPANEL.xlsx")
262 qic(doorpanel$Rejeições, chart = "pp", n=doorpanel$Amostra)
263
```

Interpretação dos resultados

A carta NP mostra o número de amostras no eixo x e o número de unidades defeituosas no eixo y. Os pontos plotados representam o número de unidades defeituosas em cada subgrupo. NP é o número médio de unidades defeituosas por subgrupo. O processo tem, em média, 1,733 unidades defeituosas para cada 25 peças produzidas.

Limites de controle

Observe que o limite inferior (LCL) é zero. A razão disso é que o número de unidades defeituosas não pode ser negativo. Pontos que caem no limite inferior não estão fora de controle.

Testes

Procure pontos localizados fora dos limites de controle ou pontos que não foram aprovados nos critérios de falta de controle. Pontos reprovados são marcados com um asterisco juntamente com o número do primeiro teste em que foi reprovado. Esta carta NP sugere que o processo de manufatura de painéis de porta está sob controle, mas produz 1,733 rejeições por cada 25 montagens.

Considerações finais

Conclusões práticas

O processo de produção de painéis de porta está sob controle, mas produz 1,733 rejeições por cada 25 montagens.

Uma vez que o processo está sob controle, pode-se realizar um estudo de capacidade para continuar a avaliação deste processo.

Considerações estatísticas

Utilize uma carta NP para avaliar se um processo está sob controle, quando o item de controle for o número de unidades defeituosas e o tamanho do subgrupo for constante. Lembre-se, “sob controle” não quer dizer que todas as unidades são conformes, mas apenas que o processo está produzindo saída de uma forma consistente.

Embora atributos possam ser mais convenientes de coletar do que variáveis, considere o seguinte :

- atributos indicam apenas se uma peça ou serviço atende às especificações; eles não indicam a extensão em que as especificações estão sendo atendidas ou não;
- atributos só são capazes de indicar se o índice de não conformidade está mudando. Variáveis contém informações sobre o centro e a dispersão da distribuição e podem indicar uma mudança no processo antes que o índice de não conformidades aumente;
- cartas de controle para atributos requerem amostras de tamanho maior do que as cartas de controle para variáveis.

Carta U

O que é uma carta U ?

A carta U monitora o número médio de não conformidades por unidade no subgrupo. Para criar esse gráfico, cada unidade é inspecionada e o número de não conformidades é registrado. O número médio de não conformidades por unidade (U) em cada subgrupo é um ponto marcado na carta U.

Para cada subgrupo, U é calculado por :

$$u = \frac{\text{nº de não conformidades}}{\text{tamanho do subgrupo}}$$

Quando usar uma carta U ?

Utilize a carta U quando estiver inspecionando à procura de não conformidades por unidade de inspeção.

Por que usar uma carta U ?

Utilize a carta U para avaliar se o processo está sob controle. A carta U pode ser usada para responder questões tais como :

- o processo de fabricação de papel de parede está sob controle no tocante ao número de defeitos ?
- o nº de erros por fatura é estável durante o processamento ?
- o nº de reclamações relacionado a serviços muda com a mudança das estações ?
- o nº de bolhas em um pára-brisa permanece sob controle ao longo da produção ?

Monitorando defeitos nos bancos Kombi

Deseja-se avaliar a qualidade do acabamento dos bancos Kombi (rugas, falhas no revestimento, falhas na estrutura, etc.). O número de defeitos de acabamento (não conformidades) para cada subgrupo está sendo contado. Uma vez que o tamanho dos subgrupos varia, pode ser usar uma carta U para plotar o número médio de não conformidades por unidade de subgrupo.

Variáveis

Os valores obtidos estão dispostos em duas colunas na planilha de

dados :

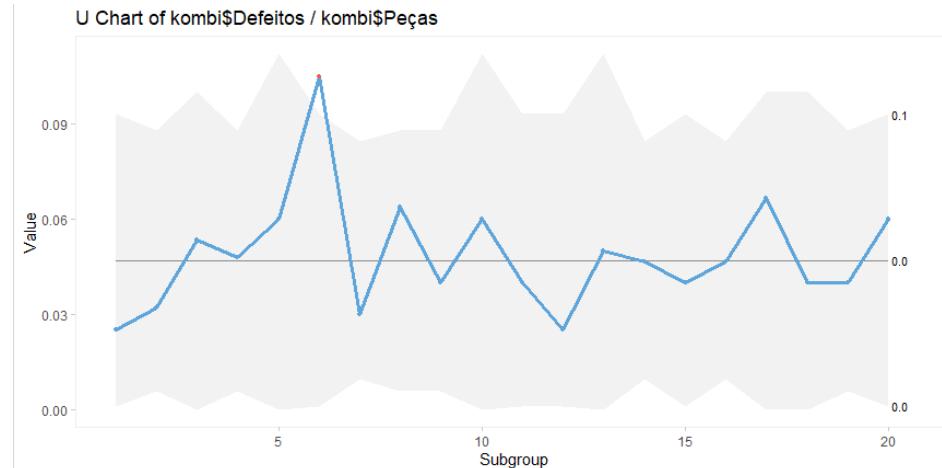
- Defeito – contém o número de não conformidades (defeitos);
- Peças – contém o número de peças avaliadas em cada subgrupo.

Testes

Quando os subgrupos não forem do mesmo tamanho, os limites de controle não serão constantes. Aqui também é possível a realização de testes à procura de causas especiais.

Carta U

```
264 #u
265 kombi<- read_excel("KOMBI.xlsx")
266 qic(kombi$Defeitos, chart = "u", n=kombi$Peças)
```



Interpretação dos resultados

A carta U mostra o número do subconjunto no eixo x e a contagem de amostras não conformes no eixo y. Os pontos plotados representam o número médio de não conformidades por unidade em cada subgrupo. U é uma estimativa da média de não conformidades por unidade.

O número médio de não conformidades é 0,0471 por banco.

Limites de controle

Uma vez que os limites de controle são em função do tamanho do subgrupo, eles mudam sempre que o tamanho do subgrupo muda. Quando o tamanho do subgrupo aumenta, os limites de controle se movem para mais perto da linha central. O MINITAB realiza testes à procura de causas especiais (veja que o subgrupo 6 está acima do limite superior e foi identificado com o teste 1).

Considerações finais

Conclusões práticas

O processo não está sob controle, conforme mostrado pelo ponto localizado acima do limite superior de controle. De acordo com estes dados, o ponto acima do limite de controle superior indica que o número médio de defeitos de acabamento aumentou mais do que o esperado para o subgrupo 6.

Deve-se buscar as causas prováveis da variação e tomar ações corretivas para evitar sua reocorrência.

Considerações estatísticas

Utilize uma carta U para avaliar a estabilidade do processo quando estiver inspecionando à procura de não conformidades por unidade de inspeção e o tamanho do subgrupo variar (você também pode usar uma carta U quando o tamanho do subgrupo é constante). Lembre-se, “sob controle” não quer dizer que todas as unidades são conformes, mas apenas que o processo está produzindo saída de uma forma consistente.

Embora atributos possam ser mais convenientes de coletar do que variáveis, considere o seguinte :

- atributos indicam apenas se uma peça ou serviço atende às especificações; eles não indicam a extensão em que as especificações estão sendo atendidas ou não;
- atributos só são capazes de indicar se o índice de não conformidade está mudando. Variáveis contém informações sobre o centro e a dispersão da distribuição e podem indicar uma mudança no processo antes que o índice de não conformidades aumente;
- atributos não são úteis quando você está tentando detectar pequenos deslocamentos de processo;
- cartas de controle para atributos requerem amostras de tamanho maior do que as cartas de controle para variáveis.

Carta C

O que é uma carta C ?

A carta C monitora o número de não conformidades no processo.

Para criar esse gráfico, cada unidade é inspecionada e o número de não conformidades encontradas no subgrupo é registrado. O número de não conformidades em cada subgrupo será um ponto marcado na carta de controle.

Quando usar uma carta C ?

Utilize a carta C quando estiver inspecionando à procura de não conformidades por unidade de inspeção.

Por que usar uma carta C ?

Utilize a carta C para avaliar se o processo está sob controle. A carta C pode ser usada para responder questões tais como :

- o nº de não conformidades em um microchip é estável ?
- o nº de defeitos em um tecido de linho permanece estável durante o processo de produção ?
- um processo de fabricação de vidro está sob controle no tocante ao número de bolhas de ar ?

Avaliando defeitos nas notas fiscais

Para ajudar a melhorar a qualidade de seus produtos, o fabricante de notas fiscais para a Lear tomou uma iniciativa de qualidade. Utilize a carta C para avaliar o número de notas fiscais não conformes (defeitos).

Variáveis

Os dados com a resposta são dispostos em uma única coluna (NF NOK), a qual contém o número de unidades defeituosas para cada subgrupo. Uma vez que o tamanho do subgrupo é constante, não é preciso uma coluna para indicar o tamanho do subgrupo.

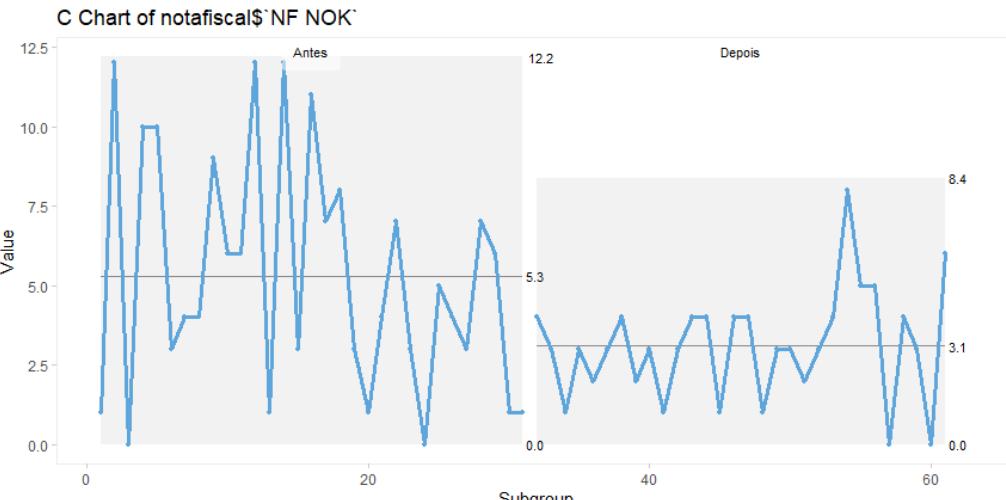
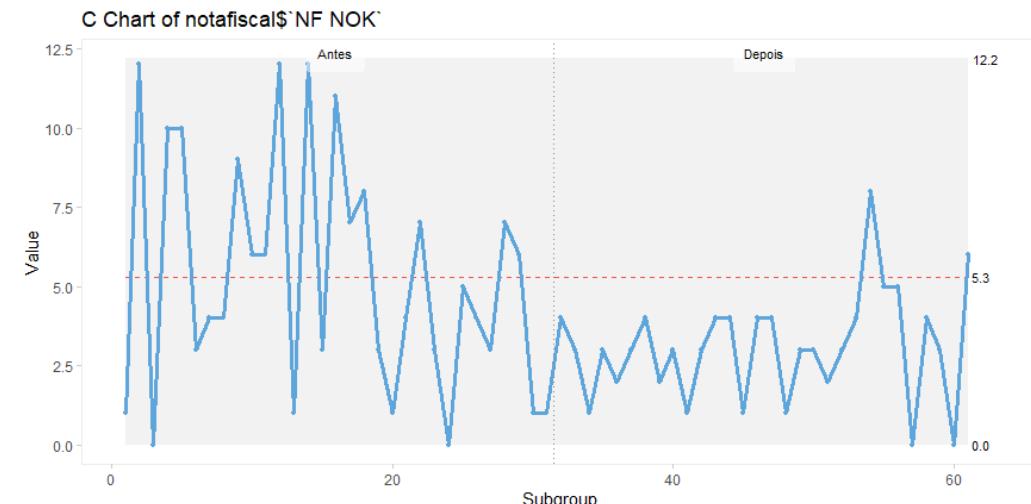
Além disso, as datas da coleta de dados são armazenadas em Data. Crie um gráfico de controle com as datas no eixo x para avaliar o sucesso da iniciativa de qualidade. A coleta de dados começou en 1º de março de 2006 e a iniciativa de qualidade começou em 1º de abril de 2006.

Testes

Quando os subgrupos são do mesmo tamanho, os limites de controle são constantes. Aqui também é possível a realização de testes à procura de causas especiais.

Carta C

```
268 #C  
269 nota fiscal<- read_excel("NOTFISC.xlsx")  
270 view(nota fiscal)  
271 qic(nota fiscal$`NF NOK`, chart = "c", freeze = 31, part.labels = c("Antes", "Depois"))  
272 qic(nota fiscal$`NF NOK`, chart = "c", part = 31, part.labels = c("Antes", "Depois"))
```



Interpretação dos resultados

A carta C mostra o número da amostra no eixo x e o número de não conformidades por subgrupo no eixo y. Os pontos plotados na carta de controle representam o número de não conformidades em cada subgrupo. C é o número médio de não conformidade no processo. Após o início da iniciativa de qualidade, o processo de impressão de notas fiscais apresentou uma média de 3,1 não conformidades por cada 50 notas fiscais

Limites de controle

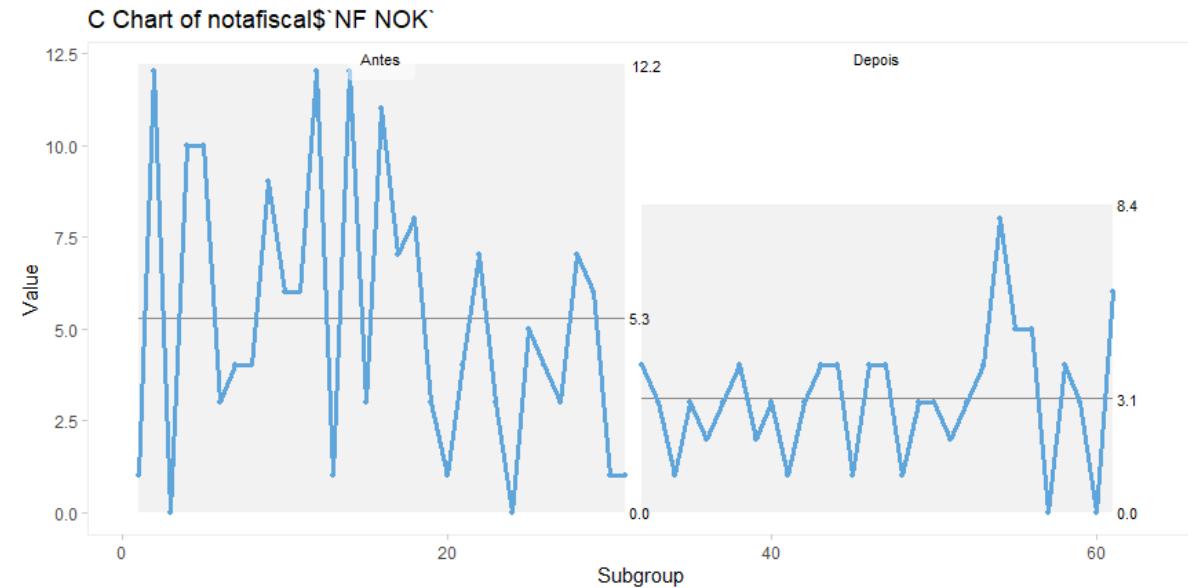
Observe que o limite inferior de controle (LCL) é zero. A razão disso é que o número de não conformidades não pode ser negativo. Pontos de dados que caem no limite inferior não estão fora de controle.

Testes

Procure por pontos localizados fora dos limites de controle ou pontos que foram reprovados segundo os critérios de falta de controle. Os pontos que foram reprovados são marcados com um ponto vermelho e o número do primeiro teste em que eles foram reprovados é indicado. Esta carta C sugere que o processo de impressão de notas fiscais está sob controle e produz 3,1 não conformidades por cada 50 notas fiscais impressas.

ISSO É BOM ?

```
268 #C
269 nota fiscal <- read_excel("NOTFISC.xlsx")
270 view(nota fiscal)
271 qic(nota fiscal$`NF NOK`, chart = "c", freeze = 31, part.labels = c("Antes", "Depois"))
272 qic(nota fiscal$`NF NOK`, chart = "c", part = 31, part.labels = c("Antes", "Depois"))
```



Agrupamento

A carta divide-se em duas seções :

- 01/03/06 a 31/03/06, que são os dados pré iniciativa
- 01/04/06 a 30/04/06, que são os dados pós iniciativa

A carta indica que o processo estava sob controle antes e depois da iniciativa. Além disso, o rebaixamento da linha central (C) sugere que a iniciativa obteve sucesso em reduzir a incidência de não conformidades. Seria bom executar um teste estatístico para verificar essa melhoria (testes de proporção, por exemplo).

Considerações finais

Conclusões práticas

O processo para fabricar notas fiscais estava sob controle antes e depois da iniciativa de melhoria do processo ser implantada.

O gráfico sequencial indica que a iniciativa obteve sucesso em reduzir a incidência de não conformidades. Conforme já dito anteriormente, seria interessante a realização de um teste estatístico para validação dessa melhoria.

Considerações estatísticas

Utilize a carta C para avaliar o controle estatístico de processo quando estiver inspecionando à procura de não conformidades e o tamanho do subgrupo for constante (você também pode usar uma carta U neste caso). Lembre-se, “sob controle” não quer dizer que todas as unidades são conformes, mas apenas que o processo está produzindo saída de uma forma consistente.

Embora atributos possam ser mais convenientes de coletar do que variáveis, considere o seguinte :

- atributos indicam apenas se uma peça ou serviço atende às especificações; eles não indicam a extensão em que as especificações estão sendo atendidas ou não;
- atributos só são capazes de indicar se o índice de não conformidade está mudando. Variáveis contém informações sobre o centro e a dispersão da distribuição e podem indicar uma mudança no processo antes que o índice de não conformidades aumente;
- atributos não são úteis quando você está tentando detectar pequenos deslocamentos de processo;
- cartas de controle para atributos requerem amostras de tamanho maior do que as cartas de controle para variáveis.



18

Análise da Tendência

Nível: Intermediário

Análise da tendência

O que é uma análise da tendência ?

O Trend Analysis (análise da tendência) é uma sistemática para analisar o comportamento do processo produtivo ao longo do tempo.

Quando usar análise da tendência ?

A análise da tendência pode ser utilizada em praticamente todos os projetos de melhoria contínua (Six Sigma, Lean, CLMC, JDI, etc.), pois facilita o entendimento sobre o comportamento do processo durante o desenvolvimento do mesmo e também após a sua conclusão.

Neste tipo de gráfico, é possível projetar valores futuros (forecast) e com isso, visualizar um possível comportamento se tudo continuar variando de acordo com os dados reais (atuais).

Por que usar análise da tendência ?

A análise da tendência pode ser usada para responder questões tais como :

- Como um processo se comporta ao longo do tempo ?
- O processo foi realmente otimizado e não existem possibilidades de instabilidade ?

Por exemplo :

- O IPPM está subindo ou caindo ao longo do tempo?
- O Yield da linha de bancos está melhorando, piorando ou está estabilizado ?
- O número de revestimentos scrapeados está diminuindo ao longo do tempo ?

Criando uma análise de tendência

O objetivo desta análise é visualizar os dados do IPPM da Planta e verificar se o mesmo está sendo reduzido após a implementação de um projeto Six Sigma.

Variáveis

- Mês – Meses em que foram coletadas os dados (de maio/04 a dezembro/05);
- IPPM – Valor do IPPM do respectivo mês.

Tipos de curvas

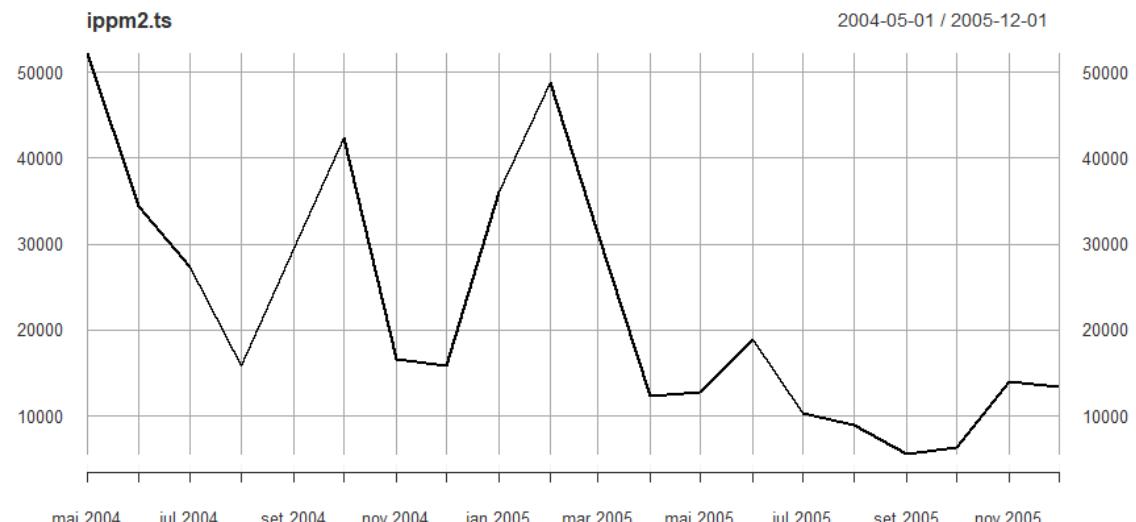
O Rstudio disponibiliza vários tipos de análise da tendência, como por exemplo :

- Linear;
- Quadrático;
- Exponencial;
- Curva-S.

Todos os tipos de análise de tendência citados acima, apresentam gráficos e equações características. Cabe ao avaliador escolher o melhor tipo, ou seja, aquele que mais se adequa às necessidades do trabalho executado.

Análise da Tendência

```
276 #Tendência
277 #install.packages("forecast")
278 library("forecast")
279 #install.packages("xts")
280 library("xts")
281 library("ggplot2")
282 #Plotando um gráfico com o timeseries
283 ippm2<- read_excel("IPPM2.xlsx")
284 ippm2$Mês <- as.Date(ippm2$Mês)
285 ippm2.ts <- xts(ippm2$IPPM, ippm2$Mês)
286 plot(ippm2.ts)
287 #Plotando um gráfico com as previsões
288 ippm2.ts <- ts(ippm2$IPPM)
289 summary(ippm2.ts)
290 ippm2.df <- data.frame(ippm = ippm2.ts, as.numeric(time(ippm2.ts)))
291 names(ippm2.df) <- c("ippm", "data")
292 plot(ippm2.ts)
293 model<-tslm(ippm~trend, ippm2.df)
294 summary(model)
295 ggplot(ippm2.df, aes(x=data, y=ippm)) + geom_point() + stat_smooth()
296 fcst<-forecast(model, h=12)
297 fcst
298 autoplot(fcst)
```



Interpretação dos resultados

Analizando os dados contidos no gráfico, podemos verificar facilmente que o IPPM está sendo reduzido ao longo do tempo.

Análise gráfica

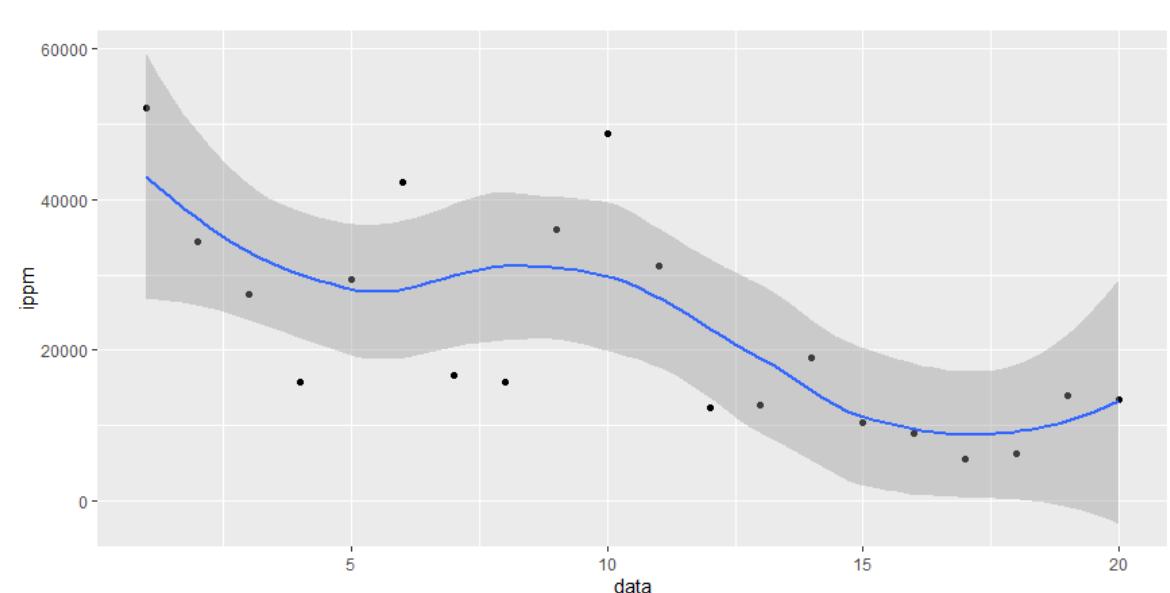
O Rstudio nos fornece algumas informações sobre o gráfico, onde podemos identificar certas características :

- Actual (linha preta) – são os dados coletados propriamente ditos;
- Fits (linha vermelha) – são os dados ajustados ao longo do tempo. É uma espécie de média teórica;
- Forecasts (linha verde) – são as projeções do processo com base nos dados ajustados ao longo do tempo.

Equação

O Rstudio nos fornece uma equação para cada tipo de análise gráfica. Neste caso, a equação fornecida é a seguinte :

```
call:  
tslm(formula = ippm ~ trend, data = ippm2.df)  
  
Residuals:  
    Min     1Q Median     3Q    Max  
-17269 -6746  -3152   6856  25402  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 39488.3    5013.3    7.877 3.06e-07 ***  
trend       -1604.1     418.5   -3.833  0.00122 **  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 10790 on 18 degrees of freedom  
Multiple R-squared:  0.4494,    Adjusted R-squared:  0.4188  
F-statistic: 14.69 on 1 and 18 DF,  p-value: 0.001219
```



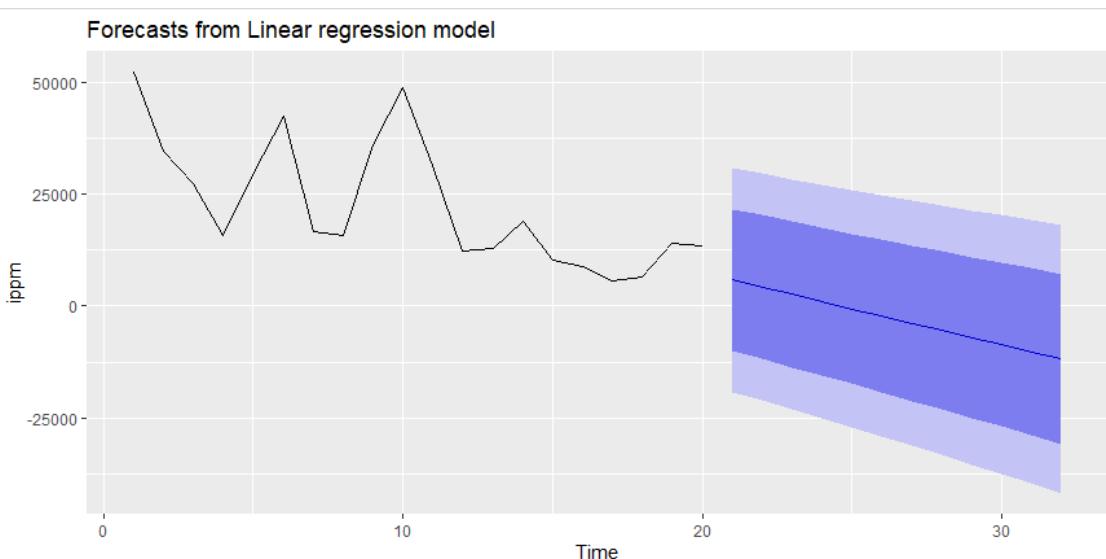
Exatidão

- **MAPE** (Mean Absolute Percentage Error ou %Erro Absoluto da Média) – mede a exatidão dos valores coletados ao longo do tempo. É expresso em porcentagem;
- **MAD** (Mean Absolute Deviation ou Desvio Absoluto da Média) – mede a exatidão dos valores coletados ao longo do tempo. Expressa a exatidão nas mesmas unidades que os dados (neste caso, IPPM), o que auxilia na definição da quantidade de erro;
- **MSD** (Mean Squared Deviation ou Desvio Quadrático da Média) – MSD é computado sempre usando o mesmo denominador (o número das projeções – Forecasts), assim você pode comparar a exatidão do valor através dos modelos e consequentemente você pode comparar a exatidão de dois modelos diferentes.

Interpretação dos resultados

Temos então os valores da previsão e podemos plotá-los com o intervalo de confiança de 80 e 95%.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
21	5801.9737	-10029.41	21633.360	-19198.58	30802.53
22	4197.8617	-11847.47	20243.195	-21140.55	29536.27
23	2593.7496	-13681.77	18869.273	-23108.17	28295.67
24	989.6376	-15531.64	17510.915	-25100.37	27079.65
25	-614.4744	-17396.38	16167.436	-27116.07	25887.12
26	-2218.5865	-19275.33	14838.155	-29154.19	24717.02
27	-3822.6985	-21167.79	13522.397	-31213.67	23568.27
28	-5426.8105	-23073.12	12219.499	-33293.45	22439.83
29	-7030.9226	-24990.66	10928.815	-35392.52	21330.67
30	-8635.0346	-26919.78	9649.715	-37509.88	20239.81
31	-10239.1466	-28859.89	8381.593	-39644.58	19166.29
32	-11843.2586	-30810.38	7123.866	-41795.70	18109.18



Considerações finais

Conclusões práticas

O IPPM está sendo reduzido ao longo do tempo. Isto mostra que o trabalho desenvolvido pelo projeto Six Sigma surtiu o efeito esperado.

A análise da tendência pode ser feita constantemente, ou seja, sem pausas, pois mesmo com os dados em projeção (Forecast), não podemos garantir com certeza que os problemas não retornarão e que com isso, o IPPM voltará a subir.

Considerações estatísticas

Podemos alterar o tipo de gráfico para uma das 4 opções existentes, e com isso analisarmos as equações disponibilizadas pelo Rstudio para o estudo.

Tanto o desvio absoluto da média quanto o desvio quadrático da média, nos fornecem informações importantes do ponto de vista estatístico, mas difíceis de interpretar do ponto de vista prático. Por isso, recomendamos a análise gráfica para tirar as principais conclusões da análise da tendência (**Trend Analysis**).



19

Glossário Linguagem Six Sigma

Nível: Básico

Glossário – Linguagem Six Sigma

A

Aleatoriedade

Uma condição na qual qualquer evento individual, em um conjunto de eventos, tem a mesma probabilidade matemática de ocorrência que todos os outros eventos do conjunto dado, isto é, os eventos individuais não são previsíveis embora eles possam pertencer, de forma coletiva, a uma distribuição definível.

Alfa (α)

É a probabilidade de se cometer Erro do Tipo I (rejeitar a hipótese nula quando ela é verdadeira) no teste de hipóteses.

Amostra

Uma ou mais observações obtidas de um conjunto maior de observações ou universo (população).

Amostra Aleatória

Uma ou mais amostras selecionadas aleatoriamente do universo (população).

Amostra Representativa

Uma amostra que reflete de forma precisa uma condição específica ou um conjunto de condições dentro do universo.

Amplitude (R)

Medida de variabilidade num grupo de dados. É calculada subtraindo-se o menor valor do grupo do maior valor do mesmo grupo.

Análise de Capacidade (Capacidade)

Análise de capacidade é uma ferramenta do programa do MINITAB que visualmente compara o desempenho do processo real aos padrões de desempenho.

Análise de Correlação

Comparar duas variáveis plotando valores paralelos. Uma variável é chamada de resposta e é plotada no eixo "Y". A outra variável, que nós estamos estudando para sua relação de resposta, é plotado no eixo "X". Freqüentemente, só plotando os pontos que se identifica se uma relação existe entre as duas variáveis. Para quantificar a relação, uma linha de regressão pode ser desenhada.

Análise de Discrepancia

Também conhecido como ANOVA. Uma técnica estatística para analisar dados experimentais. É um teste estatístico que permite a comparações de variação de origem múltipla, ou efeitos, para determinar se qualquer destas origens afeta significativamente a variabilidade do resultado estudado. Existem três modelos; fixos, aleatórios e misturados. Subdivide a variação total de um conjunto de dados em partes de componentes significantes associados com origens específicas de variação a fim de testar uma hipótese nos parâmetros do modelo ou estimar componentes de variância.

Análise de Modo e Efeitos de Falha Potencial (FMEA)

Ferramenta usada para entender os tipos potenciais de falhas, o que pode causar as falhas e como prevenir ou conter as consequências das falhas possíveis. FMEA é um procedimento utilizado para identificar modos potenciais de falha e os efeitos resultantes, de forma a avaliar a causa raiz e implantar medidas para conter ou eliminar as causas raízes das falhas.

Análise de Pareto

Método que ajuda a classificar e priorizar os problemas em duas classes: os poucos problemas vitais e os muitos triviais. Consta de cinco etapas: identificação do problema, estratificação, levantamento de dados, elaboração do Gráfico de Pareto (ver) e priorização. Na análise de Pareto são utilizadas três das Sete Ferramentas do Controle da Qualidade: Estratificação, Folha de Verificação e Gráfico de Pareto.

Análise de Processo

Método para identificar o relacionamento entre as principais causas e seus efeitos. Pode ser utilizada para identificar a causa fundamental de um problema, para se determinar padrões de valores ótimos para as causas de modo a se obter os melhores efeitos, para reduzir a dispersão, etc. Consiste em partir de um resultado e procurar a causa fundamental entre várias que compõem o processo.

Análise de Regressão

Ferramenta estatística que processa as informações contidas nos dados de forma a gerar um modelo que represente o relacionamento existente entre as diversas variáveis de um processo, permitindo a determinação quantitativa das causas mais influentes no problema ou para o alcance de uma meta.

Glossário – Linguagem Six Sigma

A

Análise de Variância

Técnica estatística que permite, no gerenciamento de processos, comparar vários grupos de interesse, mantendo um controle dos erros que podem ser cometidos no estabelecimento das conclusões.

Anomalia

Qualquer desvio das condições normais de operação. É tudo que for diferente do usual ou anormal, exigindo ação (ou atuação) corretiva.

Assimetria

Se um conjunto de dados for dividido em duas partes a partir da mediana e estas duas partes não forem iguais, então ele é dito assimétrico. Outra maneira de verificar se um conjunto é assimétrico é calcular o seu coeficiente de assimetria ou o momento de terceira ordem. Se ele for diferente de zero então o conjunto é dito assimétrico.

Atributo

Uma característica que pode assumir apenas um valor, por exemplo 0 ou 1.

B

Benchmark

Termo que indica "o(s) melhor(es) resultado(s) do mundo" dentre as organizações concorrentes, em determinados itens de controle. Observa-se atualmente a tendência de se ir além da análise dos concorrentes e procurar apresentar um desempenho ainda melhor como referencial.

Benchmarking

Processo de comparação sistemática de produtos e serviços com os oferecidos pela concorrência ou por empresas consideradas excelentes em algo determinado. O objetivo do "benchmarking" é o de conhecer e, se possível, de incorporar o que os outros estão fazendo de melhor. Tipos de Benchmarking: Interno, quando se compara atividades semelhantes dentro da mesma organização; Competitivo, quando se compara atividades semelhantes com os concorrentes; Funcional, quando se compara atividades semelhantes conduzidas dentro de empresas de ramos diferentes.

Black Belt

O líder da equipe responsável pela aplicação do processo Six Sigma. São os líderes de equipe, que atuam como agentes de mudanças, responsáveis por disseminar o conhecimento Seis Sigma e pelos projetos no processo. Eles são membros ativos da equipe e estão encarregados de coordenar as atividades da equipe, apontando as responsabilidades de cada membro, conduzindo os projetos.

Bloqueio (Bloco)

Bloqueio é um método de segregar origens conhecidas de variação, como mudanças nas pessoas, materiais ou equipamentos, e isso acontece em um experimento (como a diferença se desloca). Por bloqueio, o que afeta esta variação pode ser isolado e quantificado. A regra geral para projetar experimentos é "bloquear o que você pode, e separe aleatoriamente o que você não pode".

Bloxplot

É um gráfico que apresenta simultaneamente várias características de um conjunto de dados: locação, dispersão, simetria ou assimetria e presença de observações discrepantes (outliers). É muito útil para a comparação de dois ou mais conjuntos de dados.

Glossário – Linguagem Six Sigma

B

Brainstorming

Procedimento utilizado para auxiliar um grupo a criar o máximo de idéias no menor tempo possível. O “brainstorming” pode ser utilizado das seguintes formas: Brainstorming não Estruturado, Brainstorming Estruturado e Brainstorming Estruturado e Programado. No Brainstorming não Estruturado, os participantes do grupo dão suas idéias à medida em que elas surgem em suas mentes. Este procedimento tem a vantagem de criar uma atmosfera mais espontânea entre os integrantes do grupo. Por outro lado, pode favorecer o risco de dominação por parte dos participantes mais extrovertidos. No Brainstorming Estruturado todas as pessoas devem dar uma idéia a cada rodada ou “passar” até que chegue a próxima vez. Este procedimento estabelece uma atmosfera de certa pressão sobre o grupo, podendo gerar eventuais dificuldades durante os trabalhos. No Brainstorming Estruturado e Programado marca-se a reunião com conhecimento prévio dos temas a serem analisados e solicita-se que cada participante leve, por escrito, suas sugestões.

C

Capacidade (Capabilidade)

1. Capacidade é o alcance total da variação inherente de um processo estável. É determinado por dados de gráficos de controle. Os gráficos de controle devem indicar estabilidade antes dos cálculos de capacidade serem feitos. Os histogramas são usados para examinar a distribuição padrão de valores individuais e verificar uma distribuição normal. Quando a análise indica um processo estável e uma distribuição normal, os índices Cp e Cpk podem ser calculados. Se análise indica uma distribuição anormal, ferramentas avançadas de estatística como a análise de PPM será exigida para determinar a capacidade. Se gráficos de controle mostram que processo não é estável, o índice Ppk pode ser calculado. 2. O volume de produtos ou serviços que os negócios podem produzir com os recursos disponíveis para isto.

Característica

Um aspecto definível ou mensurável de um processo, produto ou variável.

Característica Crítica para a Qualidade - CTQ (Critical To Quality)

Uma característica do produto ou processo que é considerada crítica em relação ao impacto sobre os requisitos do cliente, desempenho, qualidade ou confiabilidade.

Características Críticas

As características críticas são aqueles requisitos de um produto (dimensões, testes de apresentação) ou parâmetros do processo que podem afetar complacentemente o regulamento do governo de assegurar a função veículo/produto e que exige fornecedor específico, assembléia, transporte, ou monitoração e inclusão no Plano de Controle. As características críticas são identificadas pelo símbolo de delta invertidos.

Caracterização Estatística de Processos

Consiste em coletar, processar e dispor as informações referentes a um processo que está sendo conduzido para manter um determinado nível de qualidade (meta padrão), com a precisão e a confiança desejadas. Trata-se de uma Avaliação de Processos (ver) mais completa, extraíndo toda a informação disponível nos dados por meio de ferramentas estatísticas de nível intermediário de complexidade.

Glossário – Linguagem Six Sigma

C

Causa Aleatória (Comum)

Uma fonte de variação que é aleatória; uma variação na fonte (variáveis "triviais") não produzirá uma variação altamente previsível na resposta (variável dependente); não existe uma correlação; qualquer fonte individual de variação resulta em uma pequena variação na resposta; não pode ser eliminada de um processo de forma econômica; uma fonte natural inerente de variação.

Causa Assinalável (Especial)

Uma fonte de variação que é não-aleatória; uma variação na fonte (variáveis "vitais") produzirá uma variação significativa, de alguma magnitude, na resposta (variável dependente), existe uma correlação; a variação pode ser devida a um efeito intermitente e em fase ou a um sistema de causa constante, que pode ser, ou não, altamente previsível; uma causa assinalável com freqüência é sinalizada por um número excessivo de pontos fora de um limite de controle e/ou por um padrão não aleatório dentro dos limites de controle; uma fonte de variação não natural; com freqüência é de interesse econômico eliminá-la.

Causa Designável

Também conhecida como uma "causa especial", este termo é usado para descrever a origem de variação em um processo que não é devido aleatório. Então pode ser identificado e eliminado.

Causalidade

O princípio de que toda mudança implica a operação de uma causa.

Causativo

Eficaz como uma causa.

Cause

O que produz um efeito ou que realiza uma mudança.

Champion

Um líder dos negócios, que proporciona uma direção de estratégia para as equipes dos projetos Seis Sigma. Ele é o responsável pelo planejamento e coordenação do programa e suporte ao Master Black Belts e Black Belts.

Ciclo PDCA

Método de controle de processos (caminho para atingir as metas estabelecidas), composto de quatro fases básicas: P (Plan) Planejamento, D (Do) Execução, C (Check) Verificação e A (Act) Ação Corretiva. Em sua forma mais simples e reduzida temos as fases: P - definição das metas e determinação dos métodos para alcançar as metas; D- educação, treinamento e execução do trabalho; C - verificação dos efeitos (resultados) do trabalho executado; A - atuação no processo em função dos resultados. Na utilização do método PDCA poderá ser preciso empregar várias Técnicas Estatísticas (ver). Conforme o tipo de metas temos: CICLO SDCA PARA MANTER (ver) ou CICLO PDCA PARA MELHORAR (ver). Também chamado Ciclo de Shewhart ou Ciclo de Deming.

Círculo de Controle da Qualidade - CCQ

Grupo voluntário e permanente de pessoas de um mesmo setor de uma organização que recebem treinamento objetivando a prática do controle de qualidade dentro deste setor, como parte das atividades do TQC (GQT). Segundo Ishikawa, K.: " As idéias básicas por trás das atividades de um CCQ são: - criação de um ambiente de trabalho feliz, no qual haja respeito à natureza do ser humano e possibilite sua satisfação; - desenvolvimento das infinitas possibilidades da capacidade mental humana e viabilização de sua aplicação; - contribuição para o melhoramento e desenvolvimento da organização" e "Não existe TQC sem CCQ. Não existe CCQ sem TQC."

Classes

Valores atribuídos a itens de uma amostra para determinar sua ocorrência relativa em uma população.

Classificação

Diferenciação de variáveis.

Classificação de Defeitos

Unidade de defeitos possíveis listados em ordem de seriedade. Comumente são usadas as classificações: classe A, classe B, classe C, classe D; ou crítico, importante, secundário e incidental; ou crítico, importante e secundário. As definições destas classificações exigem preparação e adaptação cuidadosa aos produtos, sendo provadas para ativar uma tarefa precisa de um defeito para a classificação adequada. Um plano de amostragem de aceitação separada é geralmente aplicado para classe de defeitos.

Glossário – Linguagem Six Sigma

C

Cliente

Todas as pessoas que você fornece instalação/serviços. Para um gerente de instalação, um cliente poderia ser um empregado pedindo mobília de escritório ergonomicamente correta; para um gerente de propriedade, o cliente poderia ser um inquilino que solicita melhor serviços de manutenção na construção.

Coeficiente de Variação

É o quociente entre o desvio padrão e a média de um conjunto de dados. É um percentual e portanto adequado para efetuar comparações entre diferentes conjuntos de dados.

Coleta de Dados

Fase de uma investigação na qual são levantados os dados necessários à análise posterior. Devido à sua importância no processo, deve-se dar especial atenção e cuidado a esta fase.

Confiabilidade

Capacidade de um produto ou serviço desempenhar, sem falhas, uma função requerida sob determinadas condições por um dado período de tempo. Termo também utilizado como característica de confiabilidade significando uma probabilidade ou taxa de sucesso. A probabilidade de que o produto irá executar ou cumprir, sem falhas, sua função planejada sob condições especificadas e por um período de tempo também especificado. Para aumentar a confiabilidade são de grande utilidade os métodos: FMEA - Análise do Modo de Falha e seus Efeitos e FTA - Análise da Árvore de Falhas.

Confiança

Grau em que um produto é operável e capaz de apresentar sua função exigida em qualquer tempo escolhido aleatoriamente (durante seu tempo operacional especificado, desde que o produto esteja disponível no começo daquele período). Confiança pode ser expressa pela relação: tempo disponível dividido por (tempo disponível + tempo exigido).

Configuração

O arranjo de blocos de funções baseadas em *software* em um controlador. A configuração determina as funções que o controlador pode desempenhar e em que ordem.

Controle

Acompanhamento e/ou atuação no processo (conjunto de causas) de maneira que os seus efeitos estejam em conformidade com padrões estabelecidos. O controle é exercido para manter os resultados (manter padrões) ou para melhorá-los (melhorar padrões). Controle equivale a administração, gerência. Controlar é monitorar os resultados e buscar as causas (meios) da impossibilidade de se atingir uma meta (fim), estabelecer contramedidas, montar um plano de ação, atuar e padronizar em caso de sucesso.

Controle da Qualidade

Conjunto de métodos e atividades operacionais adotadas com três objetivos principais: planejar a qualidade, manter a qualidade e melhorar a qualidade (Trilogia de Juran). Praticar o controle da qualidade é gerenciar os processos de forma a mantê-los sob controle, atuando-se na eliminação e bloqueio da causa fundamental dos problemas.

Controle de Processo

Consiste em três ações fundamentais: Estabelecimento da Diretriz de Controle; - Manutenção do nível de controle e - Melhoria do nível de controle. Controlar um Processo significa manter estável (Rotina) e melhorar (Melhorias) um conjunto de causas que afetam os itens de controle da área a ser gerenciada. É também a avaliação sistemática do desempenho de um processo e a tomada das medidas corretivas necessárias.

Controle Estatístico

Uma condição quantitativa que descreve um processo que é isento de causas assinaláveis/especiais de variação; por exemplo, variação na tendência central e variância. Uma tal condição é quase sempre evidenciada em um gráfico de controle, isto é, um gráfico de controle que mostra a ausência de variação não aleatória.

Controle Estatístico de Processo (CEP)

A aplicação de métodos e procedimentos estatísticos em relação a um processo e a um certo conjunto de padrões.

Glossário – Linguagem Six Sigma

C

Correlação

Um termo geral utilizado para descrever o fato de que duas ou mais variáveis (conjuntos de dados) estão relacionados. Galton, em 1869, foi provavelmente o primeiro a utilizar o termo com este sentido. O termo é utilizado mais precisamente para denominar relacionamento linear entre dois conjuntos de dados ou variáveis.

Covariável

Uma variável que apresenta um efeito que não se tem interesse direto. A análise da variável de interesse apresentará melhores resultados se a variação da covariável for controlada.

Custo da Qualidade

É o custo decorrente das imperfeições do processo, isto é, de não conseguir projetar, produzir e entregar 100% de qualidade aos clientes. Este custo seria eliminado se produzíssemos um produto sem falhas.

Custo de Estimativa

Uma das quatro categorias dos Custos de Qualidade Baixa. Um termo usado por alguns para descrever o custo envolvido em assegurar que alguns negócios está continuamente se esforçando em ajustar com os requisitos de qualidade dos seus clientes.

D

Dados

Informações factuais usadas como base para raciocínio, discussão e cálculos; com freqüência se trata de informações quantitativas.

Dados Contínuos (Variáveis)

Informação numérica no intervalo de nível de relação; a subdivisão tem sentido do ponto de vista de conceito; pode assumir qualquer número dentro de um intervalo, por exemplo, 14,652 amp.

Dados de Atributos

Informação numérica no nível nominal; subdivisão não tem sentido conceitualmente; dados que representam a freqüência de ocorrência dentro de alguma categoria discreta, por exemplo, 42 curtos em soldagem.

Defeito

Descumprimento de requisitos previstos para o uso do produto. É a falta de conformidade que se observa em um produto quando determinada característica da qualidade é comparada com suas especificações.

Defeito Por Milhões de Oportunidades - DPMO

Um cálculo usado nas iniciativas Seis Sigma de Melhoria do Processo que indica a quantia de defeitos em um processo por um milhão de oportunidades; isto é, o número de defeitos divididos por (o número de unidades vezes o número de oportunidades) = DPO, vezes 1 milhões = DPMO.

Design for Six Sigma (DFSS)

É o uso sistemático de ferramentas, métodos e medidas para habilitar-nos a projetar produtos e processos que atendam os níveis de qualidade Seis Sigma e a expectativa do cliente.

Desvio

A diferença entre o valor de uma variável específica e um certo valor almejado, usualmente o ponto fixado de um processo.

Desvio Padrão (s ; σ)

Um índice estatístico de variabilidade que descreve a dispersão. É a média quadrática dos desvios em relação à média aritmética.

Glossário – Linguagem Six Sigma

D

DFMEA (Design Failure Mode and Effects Analysis)

Análise de Efeitos e Modo de Defeito de Projeto.

Diagrama de Causa e Efeito (Ishikawa)

Mostra a relação entre um conjunto de causas (processo) que provoca um ou mais efeitos. É uma forma organizada de correlacionar o efeito com suas causas, agrupando-as em "famílias de causas", tais como: Matéria-Prima, Máquina, Mão de Obra, Medida, Método e Meio Ambiente. O Diagrama de Causa e Efeito proporciona ao gerente melhor entendimento de que ele tem autoridade sobre as causas e responsabilidade sobre os efeitos (resultados) de um processo. É também conhecido como Diagrama de Ishikawa ou de Espinha de Peixe. É uma das Sete Ferramentas da Qualidade.

Diagrama de Pareto

Um gráfico que classifica ou ordena ocorrências comuns.

Dispersão

O quanto um conjunto de dados está espalhado. A dispersão é normalmente avaliada em torno da média, através da variância, do desvio padrão e do desvio médio. Mas também pode ser definida pela amplitude que é a diferença entre o máximo e o mínimo do conjunto ou ainda pela amplitude inter-quartílica, isto é, a diferença entre o terceiro e o primeiro quartil.

Dispersão do Processo

A faixa de valores que uma dada característica de processo apresenta; este termo específico quase sempre se aplica à faixa mas também pode incluir a variância. A dispersão pode ser baseada em um conjunto de dados obtidos em um dado momento ou pode refletir a variabilidade ao longo de um dado período de tempo.

Distribuição Cumulativa

Uma distribuição de freqüência cumulativa é um plano do número de observações de falhas num intervalo. Uma distribuição de freqüência cumulativa pode mostrar as freqüências reais ou a porcentagem das pontuações em ou abaixo de cada intervalo. O plano pode ser um histograma ou um polígono.

Distribuição de Freqüência

O padrão ou forma apresentado pelo grupo de medidas em uma distribuição.

Distribuição Normal

Uma função densidade contínua e simétrica, caracterizada por uma curva com forma de sino. Exemplo: a distribuição de médias de amostragens.

Distribuições

Tendência de grandes números de observações se gruparem em torno de um dado valor central, com uma certa quantidade de variação ou "dispersão" para ambos os lados.

DMAIC

O DMAIC, que é o "modelo para melhorias" utilizado por um grande número de empresas que estão adotando o Programa Seis Sigma nos Estados Unidos, é um ciclo constituído por cinco fases básicas: Define (Definir); Measure (Medir); Analyze (Analizar); Improve (Melhorar); Control (Controlar). Como outros "modelos para melhorias", o DMAIC é baseado no Ciclo PDCA. Portanto, o DMAIC, assim como o Ciclo PDCA, é um método de solução de problemas ao qual é integrado, de forma lógica, um conjunto de ferramentas para coleta, processamento e disposição das informações necessárias para a execução de cada fase do método.

DPMO

Defeitos por milhão de oportunidades, é o número total de defeitos que esperamos encontrar quando produzimos um milhão de unidades. É calculado multiplicando-se o dpo por um milhão. É uma medida direta do nível sigma.

DPO

Defeitos por unidade de oportunidade, é o número de defeitos por unidade (dpu) dividido pelo total de oportunidades.

DPU

Defeitos por unidade ou o número total de defeitos esperados em uma unidade de produto. É calculado pela divisão do número total de defeitos encontrados pelo número total de unidades avaliadas.

Glossário – Linguagem Six Sigma

E

Efeito

Aquilo que foi produzido por uma causa.

Efeito Combinado

Em algumas experiências, nós podemos achar que a diferença no resultado medido correspondente ao nível de um fator não é o mesmo a todos os níveis dos outros fatores. Quando isto acontece, existe um efeito combinado entre os dois fatores. Isto também é chamado de uma interação.

Eficiência

Conjunto de atributos que evidenciam o relacionamento entre o nível de desempenho do software e a quantidade de recursos usados, sob condições estabelecidas. [NBR 13596] Tem como subcaracterísticas: comportamento em relação ao tempo e comportamento em relação aos recursos.

Erro

A variação que não prestou contas pelos efeitos estudados de uma experiência projetada ou análise de regressão que não é explicada e é chamada de "o erro".

Erro do tipo I (α)

No teste de hipóteses consiste em rejeitar a hipótese nula quando ela é verdadeira.

Erro do tipo II (β)

No teste de hipóteses consiste em aceitar a hipótese nula quando ela é falsa.

Erro Experimental

Variação nas observações feitas sob condições idênticas de ensaio. Também chamado erro residual. A quantidade de variação que não pode ser atribuída às variáveis incluídas no experimento.

Erro Quadrado Médio

É a soma dos quadrados dos desvios entre os valores do estimador (variável) e o parâmetro que ele se propõe a estimar.

Escala Sigma

O Programa Seis Sigma utiliza algumas medidas ou métricas para quantificar como os resultados de uma empresa podem ser classificados, no que diz respeito à variabilidade e geração de defeitos ou erros. A escala sigma é uma destas medidas. Na utilização da escala sigma devemos ter cuidado para que seja feita a identificação correta do número total de oportunidades de defeito. Se este número for muito grande e não corresponder à realidade, o resultado será uma falsa posição elevada na escala sigma.

Especificações de Controle

Especificações do produto que está sendo fabricado.

Estatística

1. A ciência que se preocupa com a organização, descrição, análise e interpretação de dados experimentais. É a ciência de tomar decisões diante de incertezas. 2. Alguma quantidade calculada de uma amostra, tal como a média, o desvio-padrão ou a amplitude.

Estatística Descritiva

Parte da Estatística que trata do resumo e da apresentação de conjuntos de dados.

Estatística Inferencial ou Indutiva

O processo de tirar conclusões sobre a natureza ou o modelo de populações a partir de amostras aleatórias retiradas destas populações.

Estatística Teste

É o valor amostral da estatística utilizada para testar um parâmetro no teste de hipóteses.

Estimação

Parte da inferência estatística que trata do processo de estimação e das propriedades dos estimadores.

Estimador

Um estimador é uma variável aleatória.

Estimativa

É o valor particular de um estimador, isto é, é o resultado de algum cálculo realizado sobre valores amostrais.

Glossário – Linguagem Six Sigma

E

Estudo de Reprodutibilidade e Repetitividade

Este estudo verifica o grau de adequação do sistema de medição, pela combinação de duas fontes de variação: repetitividade e reprodutibilidade. A repetitividade está associada a variações no método de avaliação (instrumento, técnica, etc.) e a reprodutibilidade está associada a variações entre operadores.

Exatidão

É a característica de uma medida que diz o quanto perto de um valor observado está do valor verdadeiro ou do valor de destino.

Excelência

Nasceu em 1982 com a publicação de *In Search of Excellence*, de Peters e Waterman. Para os autores, as empresas excelentes têm oito características distintivas: inclinação para a ação; proximidade do cliente; autonomia individual; apostar nas pessoas; criação de valores; manter-se no que se domina; simplicidade formal; e existência em simultâneo de rigidez e flexibilidade. Mais importante do que a seleção das empresas excelentes (a maioria deixou de ser alguns anos depois) e dos seus oito atributos (parte deles deixaram de ser respeitados por essas mesmas empresas), a obra foi o símbolo da nova forma de encarar a gestão.

Experimento

Um ensaio sob condições definidas, para determinar um efeito desconhecido; para ilustrar ou verificar uma lei conhecida; para testar ou estabelecer uma hipótese.

F

Facilitador

Uma pessoa especificamente treinada que funciona como um professor, treinador e moderador de um grupo, time ou organização.

Falha

A inabilidade de um item, produto ou serviço para desempenhar funções exigidas em uma demanda devido a um ou mais defeitos.

Fator

Um fator é uma denominação alternativa, utilizada na análise de variância, para uma variável. Assim se num determinado experimento consiste em determinar o efeito da dosagem de um remédio, a "dosagem do remédio" é um fator. Se neste experimento for levado em conta não apenas a dosagem mas também o tempo em que o remédio será tomado, então "o tempo" será um segundo fator. Cada valor que o fator (variável) assume é denominado nível do fator. Assim se a dosagem testada for de: 100, 110, 115 e 120 mg, este fator terá quatro níveis. Um fator deve ter pelo menos dois níveis.

Fatorial Completo

Vários fatores são investigados a vários níveis executando todas as combinações de fatores e níveis em um experimento projetado de fatorial completo. É apropriado para usar este tipo de projeto de experiência quando vários fatores forem investigados às dois ou mais níveis e interações de fatores pode ser importante.

Fatorial Fracionário

Vários fatores são investigados à vários níveis mas só um subconjunto do fatorial completo (todas as combinações possíveis) é executado no fatorial fracionário do experimento projetado. É apropriado para usar este tipo de projeto de experiência quando existir muitos fatores e níveis e é impraticável de executar todas as combinações.

Flutuações

Variâncias nos dados, causadas por um grande número de minúsculas variações ou diferenças.

Fluxo

O movimento de material em qualquer direção.

Glossário – Linguagem Six Sigma

F

Fluxograma

Diagrama que apresenta o fluxo ou seqüência normal de um trabalho ou processo, através de uma simbologia própria. Maneira gráfica de se visualizar as etapas de um processo. É o modo mais eficaz e seguro de se identificar os clientes. As principais vantagens dos fluxogramas são as seguintes: permitem uma melhor compreensão do conjunto, de modo que cada integrante da equipe tenha um pleno conhecimento do processo como um todo e não apenas do seu departamento; tornam o processo mais claro para os que não fazem parte da equipe; facilitam a fixação dos limites de cada processo ou tarefa.

FMEA - Failure Modes and Effects Analysis

Análise dos Modos de Falhas e seus Efeitos - Método de análise de produtos e processos que permite uma avaliação sistemática e padronizada de possíveis falhas, estabelecendo suas consequências e orientando a adoção de medidas corretivas (preventivas). Fornece pistas para a execução de melhorias nos sistemas, mediante a descoberta de pontos problemáticos. O ponto problemático é enfocado a partir da causa, raciocinando na direção do efeito (falha-problema). Seus objetivos são: - identificar as falhas críticas em cada componente, suas causas e consequências; - hierarquizar as falhas e analisar a confiabilidade do sistema.

Freqüência Absoluta

Número de pontos de dados que se encaixam numa determinada classe de distribuição de freqüência.

Freqüência Esperada

Número de vezes que um valor da variável deve se repetir se a hipótese nula for verdadeira ou ainda número de vezes que um valor da variável deve se repetir de acordo com um determinado modelo.

Freqüência Observada

Número de vezes que o valor de uma variável se repete.

G

Gráfico de Barras

Um Gráfico de Barras é qualquer gráfico de um conjunto de dados, como o de números de elementos de dados que caem dentro de uma ou mais categorias, que é indicado usando um retângulo cuja altura ou largura é uma função do número de elementos.

Gráfico de Controle

Uma apresentação gráfica do desempenho de uma característica ao longo do tempo em relação aos seus limites naturais e tendência central.

Gráfico de Controle de Processo

Qualquer dos diferentes tipos de gráfico nos quais os dados são lançados em função de limites de controle específicos.

Gráfico de Pareto

Gráfico de barras que ordena as frequências das ocorrências, da maior para a menor, permitindo a priorização dos problemas. Mostra ainda a curva de percentagens acumuladas. Sua maior utilidade é a de permitir uma fácil visualização e identificação das causas ou problemas mais importantes, possibilitando a concentração de esforços sobre os mesmos. É uma das Sete Ferramentas da Qualidade.

Gráficos de Linha

Gráficos usados para acompanhar o desempenho sem relação com a capacidade do processo ou com limites de controle.

Gráficos em C

Gráficos que mostram o número de defeitos por amostra.

Gráficos P

Gráficos usados para plotar a porcentagem de defeituosos em uma amostra.

Gráficos R

Plotagem da diferença entre o valor mais alto e o valor mais baixo em uma amostra. Gráfico de controle de faixa.

Gráficos X-R

Um gráfico de controle que é uma representação da capacidade do processo ao longo do tempo; apresenta a variabilidade da média e da faixa do processo ao longo do tempo.

Glossário – Linguagem Six Sigma

G

Graus de Liberdade

É a quantidade informações (variáveis) livres que serão utilizadas para o cálculo de uma estatística (fórmula). O número de valores independentes que serão utilizados na estimativa de um parâmetro. Em geral, o número de graus de liberdade de uma estimativa é igual ao número de valores utilizados no seu cálculo menos o número de parâmetros estimados no cálculo intermediário para a sua obtenção. Assim para calcular a média de uma amostra de tamanho "n", são necessários as "n" observações fazendo com que esta estatística tenha "n" graus de liberdade. Já a estimativa da variância através de uma amostra de tamanho "n" terá "n - 1" graus de liberdade, pois para a obtenção da variância amostral é necessário antes o cálculo da média amostral.

Green Belt

Uma pessoa que apoia a implementação e a aplicação de ferramentas *Six Sigma* mediante a participação em equipes de projeto.

H

Hipótese Alternativa (Ha)

Uma explicação tentativa que indica que um evento não segue uma distribuição aleatória; o contrário de hipótese nula.

Hipótese Nula (Ho)

Uma explicação tentativa que indica que uma distribuição aleatória está ocorrendo; o contrário de hipótese alternativa.

Histograma

Apresentação vertical de uma distribuição de população no que se refere a freqüências; um método formal de plotar uma distribuição de freqüência.

Glossário – Linguagem Six Sigma

I

Inferência Estatística

A utilização de amostras de uma população com o objetivo de tomar decisões sobre a população.

Instabilidade

Flutuações anormalmente grandes em um padrão.

Instrumento

Um dispositivo para medir o valor de um atributo observável; o dispositivo pode simplesmente indicar o valor observado e pode também registrá-lo ou controlá-lo.

Interação

A tendência de duas ou mais variáveis produzem um efeito combinado que nenhuma das variáveis produziria se atuasse isoladamente.

Intervalo

Categorias numéricas com unidades iguais de medida mas sem um ponto de zero absoluto, isto é, escala ou índice de qualidade.

Intervalo de Classe

Um dos alcances em que os dados de uma tabela de distribuição de freqüência (ou histograma) se fundem. Os fins de um intervalo de classe são chamados limites de classe e o meio de um intervalo é chamado uma marca de classe.

Intervalo de Confiança (CI)

O intervalo de confiança é o alcance de valores que incluem o valor verdadeiro da base de parâmetro populacional em uma probabilidade pré-atribuída chamada o nível de confiança. É a medida da certeza da forma da linha de regressão ajustada; uma faixa de 95% de confiança implica em 95% de chance que a linha de regressão verdadeira monta dentro das faixas de confiança.

L

Limite

O ponto inicial ou final de um processo que será o enfoque de um esforço da melhoria do processo.

Limite Inferior de Controle (LSL)

Uma linha pontilhada horizontal lançada em um gráfico de controle para representar o limite inferior da capacidade de um processo.

Limite Superior de Controle (USL)

Uma linha horizontal em um gráfico de controle (normalmente pontilhada) que representa os limites superiores da capacidade do processo.

Limites de Confiança

Os dois valores que definem o intervalo de confiança.

Linha Central (CL)

A linha em um gráfico de controle estatístico de processo que representa a tendência central da característica.

Glossário – Linguagem Six Sigma

M

Mapeamento do processo

Descrição gráfica, através de fluxogramas, do funcionamento de um processo. O mapeamento auxilia na visualização funcional do processo e relacionamento de suas variáveis.

Master Black Belt

Liderança qualificada, especialista em métodos quantitativos que é responsável pela estratégia, treinamento e desenvolvimento dos Black Belts. Fornece auxílio no uso de ferramentas estatísticas e de qualidade adequadas. Atua como especialista, em tempo integral, em métodos estatísticos e de qualidade trabalhando como consultor interno.

Matriz de Causa e Efeito

Uma ferramenta que permite a visualização, em forma matricial, da influência de cada parâmetro do processo (variável independente) em cada parâmetro do produto (variável dependente). Auxilia na priorização da importância de variáveis chaves de entrada no processo.

Média (X)

Modo aritmético de realçar dados. Fornece uma medida da tendência principal na distribuição de freqüência de dados. As outras medidas de tendência principal são as medianas e o modo.

Média de Processo

A tendência central de uma dada característica de processo ao longo de um determinado período de tempo ou em um momento específico.

Melhor Prática

Normalmente reconhecida como “melhor” por outras organizações, isto é um método superior ou prática inovadora que contribui para um melhor desempenho de uma organização.

Melhoria Contínua

Também chamada Melhoria Ininterrupta ou CI. É a melhoria contínua de produtos, serviços ou processos através de aumentos e melhorias de inovação.

Método Taguchi

Conjunto de técnicas estatísticas desenvolvidas por Genichi Taguchi (consultor japonês), para otimizar o projeto e a produção. Estas técnicas são usadas na elaboração de Projetos Robustos.

MINITAB

O Software Estatístico MINITAB, atualmente na versão 14, é o líder mundial neste segmento. Ele possui um conjunto de métodos que possibilita a investigação dos dados por meio de ferramentas estatísticas, facilitando a interpretação dos resultados e a tomada de conclusões precisas. O MINITAB não é só referência em Seis Sigma. Ele também é recomendado para qualquer estudo ou trabalho que envolva interpretação de dados por meio de tabelas e gráficos em várias áreas do conhecimento humano.

Modelo cliente-fornecedor

Um modelo que esboça as entradas de dados que fluem em um processo do trabalho que por sua vez, adiciona valor e produz entrega de soluções ao cliente. Também chamado Metodologia Cliente-Fornecedor.

Mudança de Cultura

Uma importante deslocamento nas atitudes, normas, sentimentos, convicções, valores, princípios operacionais e comportamento de uma organização.

Glossário – Linguagem Six Sigma

N

Não-conformidade

Uma condição dentro de uma unidade que não apresenta conformidade com uma dada especificação, padrão e/ou exigência; com freqüência chamada de “defeito”; qualquer unidade não-conforme pode ter um potencial para mais de uma não-conformidade.

Nível Aceitável de Qualidade

Nível de qualidade que, com a finalidade de inspecionar a amostragem, é o limite mediano satisfatório do processo em uma série contínua de muitos.

Nível de Confiança

É a probabilidade de que um intervalo de confiança contenha o valor do parâmetro que ele se propõe a estimar.

Nível de Significância

É a probabilidade de se cometer erro do tipo I no teste de hipóteses, isto é, a probabilidade de se rejeitar a hipótese nula quando ela é verdadeira.

Nível Sigma ou Nível de Qualidade Sigma

O nível sigma representa o desempenho do processo, isto é, a sua capacidade em atender às especificações. O nível de Qualidade Seis Sigma representa um desempenho de 99,99966% de conformidade ou 3,4 ppm de não-conformidades.

Nominal

Categorias não ordenadas que indicam a associação ou a não associação, sem qualquer implicação de quantidade. Exemplos: área de montagem número um, números de peças, etc.

Número de Aceitação

Número de máximo de defeitos ou defeitos permitidos em uma amostragem para que ela seja aceitável.

O

Objetivo

Enunciado escrito sobre resultados a serem alcançados num determinado período ou em determinada atividade. Direção a ser seguida, ponto a ser alcançado, alvo que se pretende atingir.

Oportunidade

Alguma chance para que uma falha possa ocorrer.

Otimização de Processos

Técnicas estatísticas importantes na fase de análise das causas mais prováveis do problema considerado no giro do PDCA para melhorar. Para sua utilização, na maioria das situações, será necessário realizar interfrências no processo, ou seja, deverão ser provocadas alterações planejadas e controladas nos fatores do processo com o objetivo de observar as mudanças correspondentes nos seus efeitos. Este procedimento irá gerar informações que serão processadas para que possamos confirmar as causas mais prováveis e determinar o sentido no qual o processo deverá ser direcionado, com o propósito de atingir a meta de melhoria.

Glossário – Linguagem Six Sigma

P

Padrão

Compromisso documentado, utilizado em comum e repetidas vezes pelas pessoas relacionadas com um determinado trabalho.

Planejamento de Experimentos (DOE - Design of Experiments)

Uma técnica estatística para planejar, executar e analisar experimentos. Planejamento de experimentos modela os efeitos principais e as interações das variáveis consideradas. O enfoque estatístico para a experimentação permite uma forma ordenada para coletar, analisar e interpretar dados para satisfazer os objetivos do estudo.

Plano de Amostragem Contínua

Na amostragem de aceitação, um plano, pretendido para aplicar um fluxo contínuo das unidades individuais do produto, isso envolve aceitação e rejeição de unidade por unidade. Ele emprega períodos alternativos de 100% inspeção e amostragem (a quantia relativa de 100% inspeção pendente na qualidade do produto submetido). Plano de Amostragem Contínua normalmente exige que cada t períodos de 100% de inspeção seja continuado até um número específico, e, das unidades consecutivamente inspecionadas são achados claros defeitos.

Plano de Amostragem de Aceitação

Plano específico que indica tamanho da amostragem e associa critérios de aceitação ou de não aceitação a ser usado.

Plano de Amostragem de Corrente

Numa amostra de aceitação, um plano em que os critérios de aceitação e rejeição se aplicam a amostragem cumulativa resultante para o lote atual e um ou mais lotes futuros.

Poder de um Experimento

A probabilidade de rejeitar a hipótese-nula quando ela é falsa e de aceitar a hipótese alternativa quando ela é verdadeira.

Poka-Yoke

Um sistema que cria dispositivos à prova de falhas, tendo como princípio a idéia de criar barreiras ou alertas para que não sejam cometidos erros, que possam gerar falhas.

População

Um grupo de itens similares do qual é retirada uma amostra. Com freqüência é chamada “universo”.

Precisão

1. O grau segundo o qual um valor indicado coincide com o valor real de uma variável medida. 2. Em instrumentação de processos, o grau de conformidade de um valor indicado com um valor padrão reconhecidamente aceito, ou valor ideal.

Prevenção

A prática de eliminar uma variação indesejável a priori (antes do fato). Exemplo: prever uma condição futura em um gráfico de controle e então aplicar uma ação corretiva antes que o evento previsto aconteça.

Princípio de Pareto

Estabelece que para todo problema existem poucos itens (ou causas) vitais e muitos triviais. As principais e maiores causas dos problemas estão concentradas em poucos itens vitais e não em muitos itens triviais.

Probabilidade (p(E))

A chance de que alguma coisa ocorra; a porcentagem ou número de ocorrências ao longo de um grande número de tentativas.

Probabilidade de um Evento

O número de eventos com sucesso dividido pelo número total de tentativas.

Problema

Um desvio em relação a um determinado padrão.

Procedimento Operacional Padrão - SOP

É o documento que expressa o planejamento do trabalho repetitivo que deve ser executado para o alcance da meta padrão. Contém: listagem dos equipamentos; peças e materiais utilizado na tarefa, incluindo-se os instrumentos de medida; padrões da qualidade; descrição dos procedimentos da tarefa por atividades críticas; condições de fabricação, de operação e pontos proibidos de cada tarefa; pontos de controle (itens de controle e características da qualidade) e os métodos de controle; relação de anomalias passíveis de ação; roteiro de inspeção periódicas dos equipamentos de produção. O mesmo que SOP - Standard Operation Procedure ou Procedimento Padrão de Operação.

Glossário – Linguagem Six Sigma

P

Processo

Conjunto de tarefas distintas, interligadas, visando cumprir uma missão. Conjunto de causas que produzem um ou mais efeitos (produto). Define-se um processo agrupando em seqüência todas as tarefas dirigidas à obtenção de um resultado, bem ou serviço. Isto equivale a dizer que um processo é constituído de pessoas, equipamentos, materiais ou insumos, métodos ou procedimentos informações do processo ou medidas, condições ambientais, combinados de modo a gerar um produto (bem ou serviço). Uma série de tarefas correlatas pode ser chamada de processo e um grupo de processos correlatos pode ser visto como um sistema. Qualquer organização ou empresa é um processo e dentro dela encontramos diversos processos de manufatura ou serviços. Um processo é controlado através dos seus efeitos.

Processo Estável

Um processo isento de causas assinaláveis; sob controle estatístico.

Projeto

Um problema, normalmente necessitando de uma ação planejada.

Projeto Experimental

Plano formal detalhando os particulares para conduzir uma experiência, como respostas, fatores, níveis, blocos, tratamentos e ferramentas que serão usadas.

Projeto para Seis Sigma

Também conhecido como DFSS, é uma metodologia sistemática utilizando ferramentas, treinamentos e medidas para capacitar a projetar produtos e processos que alcancem as expectativas do cliente e podem ser produzido no nível Seis Sigma de qualidade. É a aplicação das ferramentas Seis Sigma para o desenvolvimento do produto e dos esforços do Projeto do Processo com a meta de “projetar em” capacidade de desempenho Seis Sigma.

Proporção

Escala numérica que tem um ponto de zero absoluto e unidades iguais de medida do começo ao fim. Exemplo: medidas de um parâmetro de *output*, como amps.

Q

Qualidade

“Produto ou serviço de qualidade é aquele que atende perfeitamente, de forma confiável, de forma acessível, de forma segura e no tempo certo, as necessidades do cliente” (Campos, V.F.). “Qualidade deve ser definida como cumprimento de requisitos” (Crosby, P.B.). A palavra Qualidade tem diversos significados, todavia podemos especificar alguns desses significados essenciais no planejamento da própria qualidade e no planejamento estratégico da empresa. No que se refere ao desempenho, a qualidade aponta para características indicadoras da satisfação do cliente frente a produtos ou serviços. Relacionada à satisfação do cliente, a palavra qualidade também se vincula a “ausência de defeitos ou falhas”. Todavia, não podemos perder de vista que um produto ou serviço sem deficiências não significa necessariamente que satisfaça o cliente, porque algum produto ou serviço concorrente pode apresentar um desempenho melhor, atraindo o cliente. Qualidade também significa adequação ao uso.

Qualidade Total

Abrange as cinco dimensões da qualidade dos produtos e das pessoas, que afetam a satisfação das necessidades dos clientes: 1 - qualidade intrínseca do produto (bem ou serviço). 2 - custo (preço). 3 - entrega/atendimento (prazo certo, local certo, quantidade certa). 4 - moral (ver): nível médio de satisfação de um grupo de pessoas que trabalham na organização. 5 - segurança do usuário do produto e das pessoas da organização.

Quartil

Valores (são três) que dividem um conjunto de dados em quatro partes iguais.

Glossário – Linguagem Six Sigma

R

Região Crítica ou de Rejeição

A área sob a distribuição amostral que é determinada de acordo com o nível de significância do teste.

Região de Aceitação de Risco Alfa

Região de valores em que a hipótese nula é aceitável.

Região de Rejeição

A região de valores para os quais a hipótese alternativa é aceita.

Replicação

Observações feitas sob condições idênticas de ensaio.

Resolução de Problemas

O processo de resolver problemas; a separação e controle das condições que geram ou facilitam a criação de sintomas indesejáveis.

Retrabalho

Ação implementada sobre um produto não-conforme de modo que ele atenda aos requisitos especificados.

Risco Alfa

A probabilidade de aceitar a proposta alternativa quando, em realidade, a hipótese nula é verdadeira.

Risco Alfa da Região de Aceitação

A região de valores para os quais a hipótese nula é aceita.

Risco Beta

A probabilidade de se aceitar a hipótese nula quando, na realidade, a hipótese alternativa é a verdadeira.

Risco do Consumidor

Probabilidade de aceitar um lote quando, na realidade, o lote deveria ter sido rejeitado (veja Risco Beta).

Risco do Produtor

Probabilidade de se rejeitar um lote quando, na realidade, o lote deveria ter sido aceito (veja Risco Alfa).

S

Seis Sigma

Sigma é uma letra do alfabeto grego. O termo "sigma" é usado para designar a distribuição ou dispersão em torno da média de qualquer processo ou procedimento. Seis Sigma implica não haver mais que 3,4 defeitos por milhão.

Série de Experimento

É um conjunto de padrão de experiência projetada onde cada fator tem um nível dado para ser testado. Dependendo do padrão da experiência pode existir muitas séries de experiência. Isto podia também ser chamado de uma combinação de experiência ou condição experimental.

Sigma (σ)

É a décima oitava letra do alfabeto grego. O termo "sigma" é usado para designar a distribuição ou dispersão sobre a média de um processo ou procedimento. Sigma é o símbolo que representa a estatística do desvio-padrão relativo a uma população.

Subgrupo

Um grupamento lógico de objetos ou eventos que apresenta apenas variações aleatórias de evento para evento; os objetos ou eventos são agrupados para criar grupos homogêneos isentos de causas assinaláveis ou especiais. Graças à variabilidade mínima dentro do grupo, qualquer mudança na tendência central ou variância do universo será refletida na variabilidade "de subgrupo a subgrupo".

Glossário – Linguagem Six Sigma

T

Tabela de Freqüência

É usado para coletar dados para exposição gráfica. Uma tabela de freqüência é construída através da divisão de pontuação de dados coletados em intervalos e contando o número de pontuações em cada intervalo. O número real de pontuações como também a porcentagem de pontuações em cada intervalo é exibida. O princípio é o mesmo para uma “tabela de registro”, embora em vez de números, registros são usados.

Técnicas Estatísticas

Na utilização do Ciclo PDCA poderá ser necessário empregar ferramentas para a coleta, o processamento e a disposição das informações necessárias à condução das várias etapas do ciclo. Estas ferramentas serão denominadas Ferramentas de Qualidade e entre elas são de especial importância as técnicas estatísticas. Algumas destas técnicas são: Sete Ferramentas do Controle da Qualidade, Amostragem, Análise de Variância, Análise de Regressão, Planejamento de Experimentos, Otimização de Processos, Análise Multivariada e Confiabilidade.

Tendência Central

Média numérica, por exemplo, média, mediana e moda; linha central em um gráfico de controle estatístico de processo.

Teoria

Um princípio geral plausível ou cientificamente aceitável apresentado para explicar fenômenos.

Teste de hipóteses

São testes de significância estatística, com o objetivo de decidir se uma afirmação em relação a um parâmetro da população é falsa ou verdadeira, ou ainda se uma afirmação em relação a um parâmetro de duas populações é falsa ou verdadeira, com base em dados amostrais.

Teste de Significância

Um procedimento para verificar se uma quantidade submetida a uma variação aleatória difere de um valor postulado por uma quantidade superior à esperada por uma variação aleatória apenas.

Teste dos sinais de Wilcoxon

Um teste não paramétrico ou de distribuição livre para testar a diferença entre duas populações utilizando amostras emparelhadas. O teste toma por base as diferenças absolutas dos pares de observações das duas amostras, ordenados de acordo com o seu valor onde cada posto (diferença) recebe o sinal da diferença original. A estatística teste é a soma dos postos positivos.

Teste não-paramétrico

Um teste não paramétrico testa associações, dependência/independência e modelos ao invés de parâmetros.

Teste t - de Student

Teste paramétrico que utiliza duas amostras independentes. Testa a diferença entre duas médias populacionais quando os desvios padrões populacionais são desconhecidos (o que ocorre na grande maioria dos casos).

Glossário – Linguagem Six Sigma

U

Unidade não-conforme

Uma unidade que não apresenta conformidade com uma ou mais especificações, padrões e/ou exigências.

Usabilidade

Conjunto de atributos que evidenciam o esforço necessário para se poder utilizar o software, bem como o julgamento individual desse uso, por um conjunto explícito ou implícito de usuários. Tem como subcaracterísticas: inteligibilidade, apreensibilidade e operacionalidade.

V

Variabilidade

Também denominada variação ou dispersão, é uma característica inerente a todo processo, segundo a qual a medição de qualquer item de controle nunca se repete com o mesmo valor. Para um conjunto de valores medidos, a variabilidade pode ser medida pela amplitude (ver), ou pelo desvio padrão (ver). É o resultado de alterações nas condições sob as quais as observações são tomadas. Estas alterações podem refletir diferenças entre as matérias-primas, as condições dos equipamentos, os métodos de trabalho, as condições ambientais e os operadores envolvidos no processo. A variabilidade é a causa da fabricação de produtos (bens ou serviços) defeituosos. As causas da variação podem ser causas comuns ou causas especiais.

Variabilidade de Causa Comum

Variabilidade de causa comum é a origem de variação causada por fatores desconhecidos que resultam em uma distribuição fixa, mas aleatória de output em torno da média de dados. A variação de causa comum é a medida do potencial do processo, ou o quanto bem o processo pode desempenhar quando a variação de causa especial é removida. Por isso, é uma medida da tecnologia do processo. A variação de causa comum também é chamada variação aleatória, ruído, variação não-controlável, variação dentro do grupo, ou variação inerente. Exemplo: muitos Xs com um choque pequeno.

Variação

Qualquer diferença quantificável entre medidas individuais; tais diferenças podem ser classificadas como provenientes de causas comuns (aleatórias) ou de causas especiais (assinaláveis).

Variações Aleatórias

Variações nos dados que resultam de causas que não podem ser identificadas ou controladas.

Variações Assinaláveis

Variações nos dados que podem ser atribuídas a causas específicas.

Variações Designáveis

Variações de dados que podem ser atribuídas à causas específicas.

Variância (s^2 ; σ^2)

É a média do quadrado das distâncias euclidianas que cada ponto do conjunto está da média aritmética.

Glossário – Linguagem Six Sigma

V

Variância Relativa

É o quociente entre a variância e o quadrado da média.

Variáveis de Bloqueio

Um conjunto relativamente homogêneo de condições dentro do qual diferentes condições das variáveis primárias são comparadas. Usadas para garantir que as variáveis de fundo não contaminarão a avaliação das variáveis primárias.

Variáveis de Controle Primário

As principais variáveis independentes usadas em um experimento.

Variáveis de Fundo

Variáveis que não apresentam interesse experimental e que não são mantidas constantes. Admite-se que seus efeitos são, com freqüência, insignificantes ou desprezíveis, ou que variam de forma aleatória de modo a assegurar que não ocorre contaminação da resposta primária.

Variáveis Sistemáticas

Um modelo que apresenta tendências previsíveis.

Variável

1. Um item de dados que pode assumir qualquer conjunto de valores. 2. Uma característica que pode assumir diferentes valores.

Variável Aleatória

Uma variável que pode assumir qualquer valor de um conjunto de valores possíveis.

Variável Aleatória Contínua

Uma variável aleatória que pode assumir qualquer valor continuamente em um dado intervalo.

Variável Aleatória Discreta

Uma variável aleatória que pode assumir apenas valores entre um número definido de valores discretos.

Variável Contínua

Uma variável contínua é uma que, dentro dos limites do alcance da variável, qualquer valor é possível. Por exemplo, a variável "Tempo para resolver uma palavra cruzada" é contínuo porque pode levar 5 minutos, 7.13 minutos etc. para terminar um problema. A variável "Número de respostas corretas um teste de múltipla escolha de 100 pontos" não é uma variável contínua porque não é possível conseguir 54.12 respostas corretas. Uma variável que não é contínua é chamada "discreta."

Variável Controlada

1. A variável que o sistema de controle procura manter no valor do ponto fixado. O ponto fixado pode ser constante ou variável. 2. A parte de um processo que se deseja controlar (fluxo, nível, temperatura, pressão, etc). 3. Uma variável de processo que se deve manter em um certo valor desejado por meio da atuação em outra variável do processo.

Variável Dependente

Uma variável de resposta; por exemplo, y é a variável dependente ou variável de "Resposta" na expressão $y = f(X_1, \dots, X_n)$.

Variável Independente

Uma variável controlada; uma variável cujo valor é independente do valor de outra variável.

Variável Manipulada

1. Em um processo que visa a regular alguma condição, uma quantidade ou uma condição que é alterada pelo controle a fim de iniciar uma mudança no valor da condição regulada. 2. A parte do processo que é ajustada para eliminar a diferença entre o ponto fixado e a variável controlada.

Variável Medida

1. A quantidade física, propriedade ou condição que deve ser medida. As variáveis comuns a medir são temperaturas, pressão, velocidade de fluxo, espessura, velocidade, etc. 2. A parte do processo que é monitorada para determinar a condição real da variável controlada.

Vício (Tendência)

Conceito referente a uma estatística ou estimador. Um estimador é dito sem vício (não-viciado), não-tendencioso ou imparcial se a média de sua distribuição amostral coincide com o parâmetro a ser estimado.

BIBLIOGRAFIA

- www.r-project.org
- www.rstudio.com