Estudo de Caso 01: Estimação de Quantidades Pecuniárias

Mariana Fonseca, Taiguara Tupinambás, Douglas da Silva e Darcílio Pereira

17 de Abril, 2017

1. Sumário

Durante a aula do dia 11/04/2017 do curso *Design and Analysis of Experiments*, cada um dos 29 alunos presentes deu sua estimativa para duas perguntas do professor Felipe Campelo:

1) Quantas moedas há no recipiente 1 (um copo de 200ml com moedas variadas padrão monetário Real)?
2) Qual o valor total contido no recipiente 2 (um copo de 200ml com moedas de R$0,05)?

No primeiro caso, as estimativas foram dadas de maneira cega (sem a interação entre alunos). No segundo caso, as estimativas foram dadas de maneira aberta (declarações sequenciais dos alunos em voz alta).

Os propósitos deste estudo de caso são, com base nas estimativas fornecidas, a) **definir a quantidade e o valor real** das perguntas acima, além de b) **investigar se conjuntos de alunos de engenharia são bons estimadores de pequenas quantidades pecuniárias**.

Para isso, inicialmente é feito um estudo de estatística descritiva do conjunto de dados coletados (as estimativas dos alunos), através de gráficos e descrições paramétricas, respondendo ao propósito **a** e obtendo uma ideia da qualidade do estimador conjunto de alunos de engenharia.

Como não se sabe, a priori, a quantidade real (ou a média real da variável aleatória *quantidade de moedas em um copo cheio de 200ml*), são propostos testes adicionais com experimentação, para aprofundar a investigação do propósito **b**, através do estudo das seguintes questões de interesse:

- As estimativas rejeitam uma hipótese nula mais próxima da real?
- As estimativas de conjuntos de alunos são consistentes entre si?
- As estimativas de conjuntos de alunos são sistematicamente rejeitadas por experimentação ?

2. Preparação e Exploração dos Dados

Os dados coletados em sala são carregados no R e em seguida explorados, para se ter uma ideia dos valores obtidos

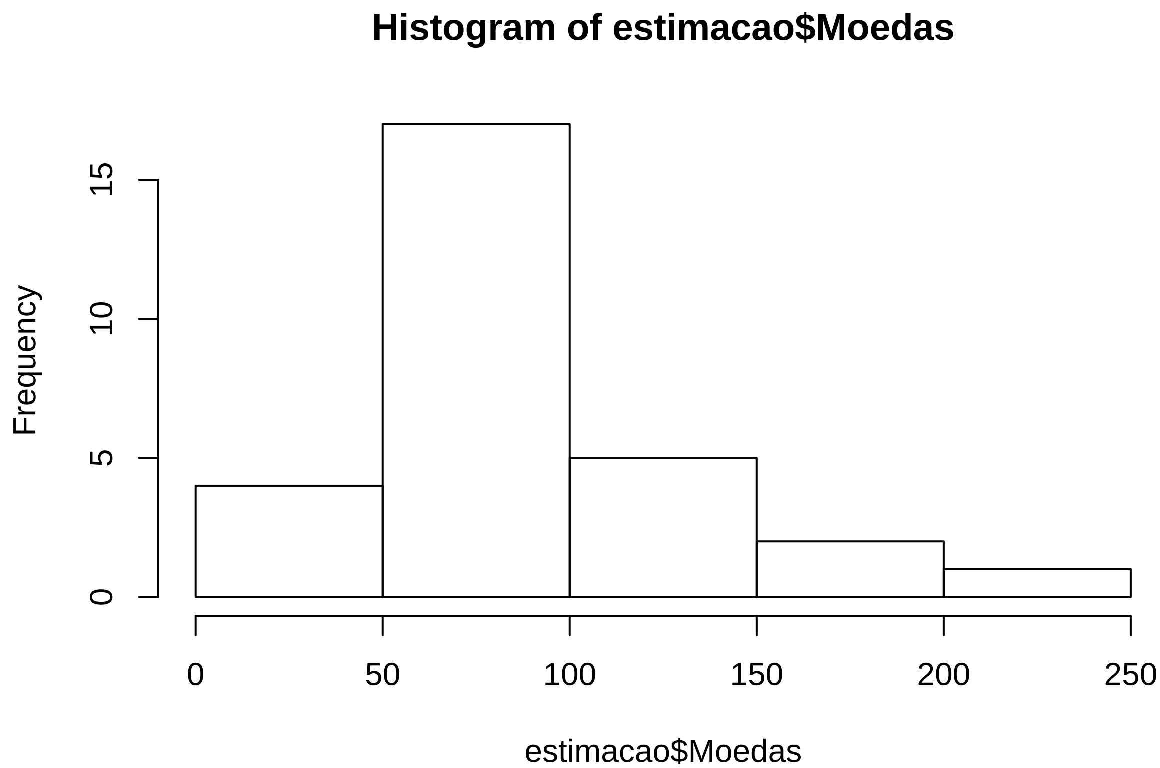
```
estimacao <- read.csv(file="./estimativa_da_sala.csv", header=TRUE, sep=";")
```

Primeiro, as estimativas de moedas são exploradas.

```
summary(estimacao$Moedas)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   34.00   57.00   67.00   88.59  107.00  250.00
```

```
hist(estimacao$Moedas)
```




```r
shapiro.test(estimacao$Moedas)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  estimacao$Moedas
## W = 0.79755, p-value = 7.551e-05
```

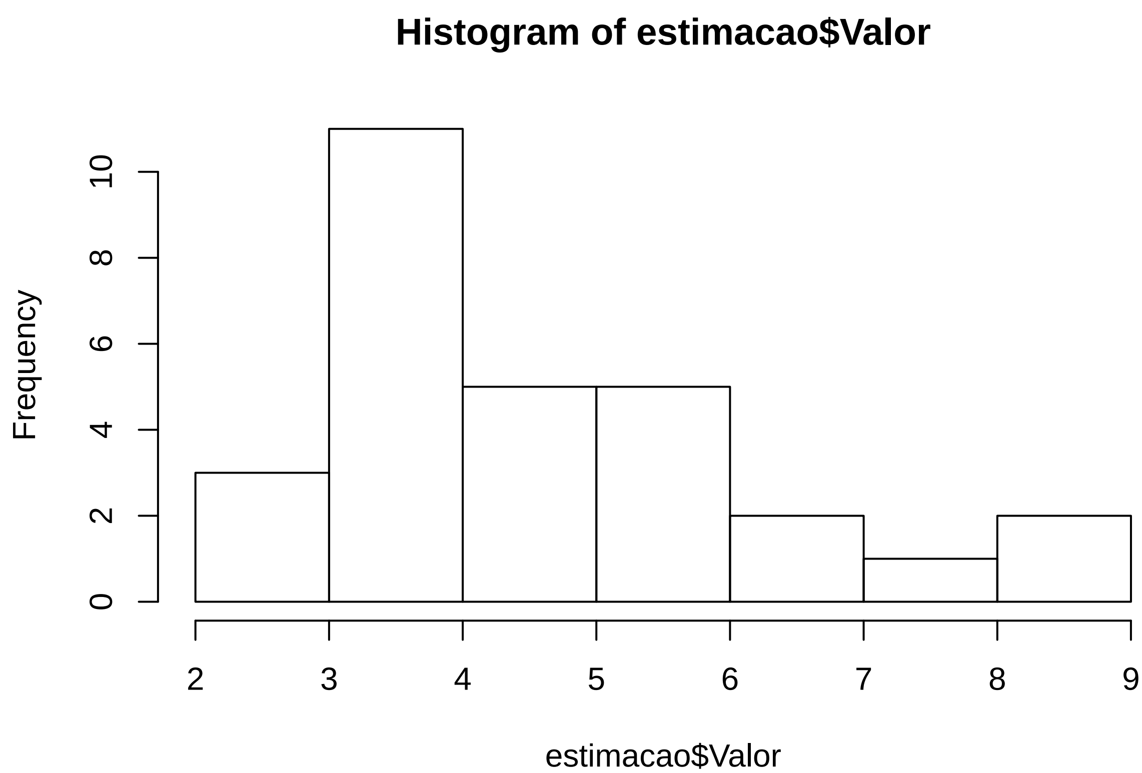
Os dados estimados para a quantidade de moedas tem uma distribuição que pode ser aproximada da normal, uma vez que o valor p<<<0.1 para o teste Shapiro-Wilk. [**1**]. Além disso, como não há a presença de valores discrepantes (*outliers*) significativos e como não se trata de valores nominais, o parâmetro do estimador será definido como a média [**2**].

Em seguida, as estimativas de valores são exploradas.

```r
summary(estimacao$Valor)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.40    3.50    4.25    4.62    5.30    8.45
```

```r
hist(estimacao$Valor)
```




```r
shapiro.test(estimacao$Valor)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  estimacao$Valor
## W = 0.91928, p-value = 0.02923
```

Assim como para as estimativas de moedas, a distribuição para as estimativas de valor pode ser assumida como normal pelo mesmo motivo, $p<0.1$, mas com um valor significativamente maior. O maior resultado de p, significando uma "menor normalidade" na distribuição pode ser explicada pela não independência dos dados, devido à forma como foram coletados (aberta, em voz alta). De qualquer forma, pelos resultados e pelas mesmas justificativas das estimativas de moedas, será utilizado o parâmetro média para o estimador.

3. Estatística Descritiva das Estimativas

3.1 Quantidade de Moedas no Recipiente 1

3.1.1 Descrição da coleta de dados

As estimativas dos alunos para a primeira pergunta foram escritas em folhas de papel e entregues ao professor, que informou aos alunos o valor de cada uma.

Assume-se que:

- Não houve interação entre os alunos (i.e. nenhum aluno trapaceou e consultou outros a respeito das estimativas ou leu do papel de um colega alguma estimativa).

- Não houve erro na leitura das estimativas por parte do professor.
- Não houve erro na anotação das estimativas por parte do grupo.

No caso da última premissa, outros grupos foram consultados e os valores conferidos, de forma a minimizar a chance de ocorrência de erro. Em relação às duas primeiras, a verificação só poderia ser feita caso o professor tenha guardado os papéis ou, e.g. houvesse uma câmera na sala para flagrar algum aluno trapaceando as instruções.

3.1.2 Análise Estatística

Após a análise exploratória dos dados, para definir a quantidade real de moedas no recipiente 1, baseado nos dados coletados em sala, é calculada a média e o erro médio das estimativas [**3**]:

$$\hat{\sigma}_X = \frac{S}{\sqrt{n}} \tag{1}$$

```
med_est_moed<-mean(estimacao$Moedas)
desvio_pad_moed<-sd(estimacao$Moedas)
erro_med_moed<-sd(estimacao$Moedas)/sqrt(length(estimacao$Moedas))
inter_conf_moed<-c(med_est_moed-2*erro_med_moed,med_est_moed+2*erro_med_moed)
```

```
## Estimativa de Média é: 88.58621
```

```
## Intervalo de Confiança 95% é: 70.22837 106.944
```

```
## Desvio padrão do estimador é: 49.42998
```

Desta forma, considerando que o estimador em questão não posssui viés, i.e.:

$$E(\Phi) = \theta \tag{2}$$

A quantidade de moedas no recipiente 1, com 95% de confiança, é de 89 $\pm$ 18. Ou seja, este intervalo tem uma chance de 95% de capturar a quantidade de moedas real no recipiente 1, considerando que ele não tenha viés e que sua distribuição seja normal.

O desvio padrão encontrado, de 49.43, indica uma dispersão muito grande em torno da estimativa de média (56% do valor da média), o que indica uma precisão muito pequena do estimador.

3.2 Valor no Recipiente 2

3.2.1 Descrição da coleta de dados

Para a segunda pergunta, as estimativas do valor contido no recipiente 2 foram dadas em voz alta, na ordem em que os alunos se encontravam e os grupos foram anotando as respostas.

Desta forma, será assumido apenas que:

- Não houve erro na anotação das estimativas por parte do grupo

Da mesma forma que a primeira estimativa, os valores anotados foram conferidos entre grupos para verificar essa premissa.

3.2.2 Análise Estatística

Após a análise exploratória dos dados, para definir o valor real de moedas no recipiente 2, baseado nos dados coletados em sala, é calculada a média e o erro médio das estimativas.

```
med_est_val<-mean(estimacao$Valor)
desvio_pad_val<-sd(estimacao$Valor)
erro_med_val<-sd(estimacao$Valor)/sqrt(length(estimacao$Valor))
inter_conf_val<-c(med_est_val-2*erro_med_val,med_est_val+2*erro_med_val)
```

```
## Estimativa de Média é: 4.619655
```

```
## Intervalo de Confiança 95% é: 4.033196 5.206114
```

```
## Desvio padrão do estimador é: 1.57909
```

Desta forma, considerando que o estimador em questão não posssui viés, o valor contido no recipiente 2, com 95% de confiança, é de R$4.60 $\pm$ R$0.60. Ou seja, este intervalo tem uma chance de 95% de capturar o valor real contido no recipiente 2, considerando que ele não tenha viés e que sua distribuição seja normal.

O desvio padrão encontrado, de R$1.58, indica uma dispersão muito grande em torno da estimativa de média (34% do valor da média), o que significa uma precisão muito pequena do estimador. No entanto, observou-se um valor percentual menor do que para a quantidade de moedas, que pode ser explicado pelo método de coleta de dados, em que cada aluno pode ter sido influenciado pela opinião dos colegas que estimaram antes dele.

3.3 Análise da Estatística Descritiva

Os estimadores para quantidade de moedas e valor em copos de 200ml produziram, neste caso, resultados pouco precisos, o que poderia indicar uma má qualidade na estimação. Além disso, a premissa de que eles não possuem viés não pode ser confirmada a priori, uma vez que o valor real dos parâmetros não são conhecidos. Para aprofundar essa investigação, novos experimentos foram propostos, através da experimentação, como detalhado nas seções seguintes.

4. Experimentos Adicionais - Recipiente 1

4.1 As estimativas rejeitam uma hipótese nula mais próxima da real?

4.1.1 Planejamento Experimental

O primeiro experimento conduzido tem como método a Prova por Contradição através do teste de hipóteses. Ou seja, através de experimentos, são criadas amostras da população de interesse, i.e. a variável aleatória *quantidade de moedas em um copo de 200ml cheio de moedas variadas.* A partir das amostras, é definida uma hipótese nula (média do resultado experimental μ_e) a ser testada pelas estimativas produzidas em sala:

$$H_0 : \mu = \mu_e \tag{3}$$
$$H_1 : \mu \neq \mu_e \tag{4}$$

A partir de parâmetros experimentais adequados, os resultados podem levar a duas conclusões:

- H_0 rejeitada: como a hipótese nula é um valor mais próximo do real, caso ela seja rejeitada, pode-se concluir que, *para este caso específico*, conjuntos de alunos de engenharia não produziram um bom estimador, uma vez que "um estimador deve ser de alguma forma próximo do valor real do parâmetro desconhecido", de acordo com a Seção 7.2.1, de [3].

- Falhou em rejeitar H_0: se, com as estimativas da sala, evidências suficientes não foram encontradas para rejeitar a hipótese nula, conclui-se que, *para este caso específico*, conjuntos de alunos de engenharia produziram um bom estimador.

É importante explicar o termo destacado em itálico, que significa que as conclusões desse teste servem apenas para qualificar ou desqualificar os estimadores para uma estimativa específica, não podendo ser generalizada. Para tentar investigar mais profundamente, novos teste serão necessários.

4.1.1.1 Desrição da coleta de dados

Um copo de 200ml semelhante ao utilizado pelo professor e diversas moedas dos variados tipos foram coletadas, de forma que a quantidade total fosse muito maior do que a capacidade do copo. Para que as amostras fossem feitas da forma mais fiel possível à real (coleta de moedas da "caixinha da paçoca" do laboratório ORCS da UFMG), a proporção de moedas de 1 real, cinquenta, vinte e cinco, dez e cinco centavos deveria ser igual (considerando que moedas de 1 centavo não foram inseridas na caixinha). Como a proporção real não tem como ser averiguada, a proporção do experimento será comparado ao meio circulante nacional do dia 11/04/2017 (data esperada da coleta de moedas), informação disponível no site do Banco Central do Brasil [4].

Para fins de reproducibilidade a quantidade de moedas de cada tipo é apresentada: 22 moedas de 1 real, 21 moedas de 50 centavos, 30 de 25 centavos, 60 de 10 centavos e 43 de 5 centavos.

A comparação das proporções de moedas do experimetno com o meio circulante nacional são apresentadas a seguir:

```
moedas_exp<-c(22,21,30,60,43)
prop_exp<-round(moedas_exp/sum(moedas_exp)*100)
prop_mcn<-c(16,13,14,30,28)
moedas<-c("1","0,50","0,25","0,10","0,05")

print(data.frame(moedas,prop_exp,prop_mcn))
```

```
##   moedas prop_exp prop_mcn
## 1      1       12       16
## 2   0,50       12       13
## 3   0,25       17       14
## 4   0,10       34       30
## 5   0,05       24       28
```

```
## Estimativa de Média é: 4.619655
```

```
## Intervalo de Confiança 95% é: 4.033196 5.206114
```

```
## Desvio padrão do estimador é: 1.57909
```

Nota-se uma diferença pequena nas proporções das moedas do experimentos com as moedas do meio circulante nacional, o que sugere que a aproximação do experimento com o real é boa.

As moedas coletadas são apresentadas na Fig. 1 e um exemplo de amostra é apresentado na Fig. 2.

As moedas foram reunidas em um estojo que era misturado antes de cada amostra. O copo era então inserido no estojo e enchido de moedas para ser contado. Ao ser removido do estojo, era feita uma avaliação pelo grupo se o nível de enchimento do copo era semelhante ao do dia das estimativas e, em caso negativo, uma nova amostra era coletada.

É importante ressaltar os pontos fracos dessas premissas:

- Caso a proporção das moedas utilizadas para encher o copo seja muito distante da utilizada no experimento ou o formato do copo seja muito diferente, as amostras observadas podem conter um viés

Figure 1: Moedas utilizadas no experimento

Figure 2: Exemplo de amostra

significativo, prejudicando o experimento. Para evitar esse problema, poderia-se verificar a proporção real ou utilizar o mesmo conjunto de moedas e copo utilizado pelo professor.
- O nível de enchimento também pode influenciar na quantidade de amostras. O grupo tentou ser o mais fiel possível, mas um erro sistemático nessa estimativa pode gerar um viés. O que poderia ter sido feito para evitar essa questão é a medição do espaço vazio do copo utilizado na sala de aula, para que a replicação fosse mais fiel.

Para as amostras aprovadas, a quantidade de moedas foi contada e anotada.

Como o objetivo desse experimento é encontrar uma hipótese nula a ser usada na prova por contradição, não é necessário um cálculo de tamanho amostral. A quantidade de amostras foi definida de forma qualitativa, observando-se a dispersão entre cada amostra.

4.1.1.2 Análise Exploratória dos Dados

Os resultados observados após 5 amostras foi considerado satisfatório (desvio padrão relativo pequeno), para se definir uma hipótese nula.

```
experimento <- read.csv(file="./resultados_experimentais.csv", header=FALSE, sep=";")
```

```
## [1] 113 114 108 119 120
```

```
## Média dos resultados experimentais é: 114.8
```

```
## Desvio padrão dos resultados experimentais é: 4.868265
```

4.1.2 Teste de Hipóteses

Para o teste de hipóteses da prova por contradição, queremos ter certeza que, no caso de rejeição da hipótese nula, a conclusão seja forte. Dessa forma, é importante minizar o erro tipo I, que seria a rejeição da hipótese nula, quando ela é verdadeira. Por isso, será considerado um nível de confiança de 99%

```
t.test(estimacao$Moedas,mu=mean(experimento$V1),conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  estimacao$Moedas
## t = -2.8559, df = 28, p-value = 0.007998
## alternative hypothesis: true mean is not equal to 114.8
## 99 percent confidence interval:
##   63.22245 113.94996
## sample estimates:
## mean of x
##  88.58621
```

De acordo com o resultado do teste, a hipótese nula deve ser rejeitada, com valor p=0.007998. Ou seja, até um nível de confiança de 99.2%, a hipótese nula seria rejeitada. Pode-se concluir por contradição que, para essa estimativa específica, o estimador conjunto de alunos de engenharia produz um parâmetro com viés.

Para que uma conclusão mais generalizada possa ser alcançada, outro teste deve ser proposto.

4.2 As estimativas de conjuntos de alunos são sistematicamente rejeitadas por experimentação?

4.2.1 Planejamento Experimental

Para que o estimador pudesse ser testado mais vezes, para uma conclusão mais generalizada pudesse ser alcançada, as estimativas dos alunos foram divididas em 5 grupos (4 grupos com 5 estimativas e 1 grupo com 4). Como as observações foram feitas de forma independente, não há necessidade de randomização das estimativas.

Nesse caso, será feita uma abordagem diferente, na qual a hipótese nula é a média de cada grupo (μ_g) e o experimento é utilizado no teste de hipóteses, ou seja:

$$H_0 : \mu = \mu_g \tag{5}$$
$$H_1 : \mu \neq \mu_g \tag{6}$$

Desta forma, é possível se ter um maior controle do experimento, uma vez que o tamanho amostral pode ser definido para se obter os parâmetros desejados.

Nesse caso, como já é esperado que a média dos grupos não seja próxima da média real, é necessário minimizar o erro de tipo II. Para isso, uma alta potência será necessária e será garantida pelo cálculo do tamanho amostral. Os parâmetros definidos são:

- Potência = 0,90 (para evitar erros do tipo II)
- $\delta = 5$ (aproximadamente o desvio padrão do experimento já realizado)
- $\sigma = 4.9$ (desvio padrão do experimento realizado)
- $\alpha = 0.1$ (não há necessidade de baixo nível de significância)

```r
power.t.test(power=0.9,delta=5,sd=4.9,sig.level=0.1,type="one.sample",alternative="two.sided")
```

```
##
##      One-sample t test power calculation
##
##               n = 9.747895
##           delta = 5
##              sd = 4.9
##       sig.level = 0.1
##           power = 0.9
##     alternative = two.sided
```

O tamanho amostral requerido para se atingir a potência desejada, considerando os parâmetros estimados é de 10, i.e., 5 testes adicionais são necessários.

A seguir os grupos são divididos e os valores das hipóteses nulas de cada um são calculados:

```r
h_0<-c(0,0,0,0,0)
aux<-0
for (i in 1:5) {
        aux<-estimacao$Moedas[((i-1)*6+1):((i-1)*6+6)]
        h_0[i]<-mean(aux,na.rm = TRUE)
}
cat("As hipóteses nulas são:",h_0)
```

```
## As hipóteses nulas são: 73.83333 92.16667 104.5 95 75.2
```

4.2.1 Testes de Hipótese

Em seguida é realizado o teste para a hipótese nula de cada grupo e os valores do intervalo de confiança, além do valor p, são guardados em uma tabela.

```
library(broom)
conf_low<-c(0,0,0,0,0)
conf_high<-c(0,0,0,0,0)
p_value<-c(0,0,0,0,0)
ttest<-0
reject<-0
for (i in 1:5) {
        ttest<-tidy(t.test(experimento$V1,mu=h_0[i],conf.level = 0.9))
        conf_low[i]<-ttest$conf.low
        conf_high[i]<-ttest$conf.high
        p_value[i]<-ttest$p.value
        if (h_0[i]<conf_low[i] | h_0[i]>conf_high[i]) {
                reject<-reject+1
        }
}
```

```
## A hipótese nula foi rejeitada 5 vezes do total de 5 grupos
```

```
## Os valores p dos testes foram 4.697341e-05 0.0004834902 0.009097731 0.0008106394 5.373114e-05
```

Como a hipótese nula (estimativa dos grupos) foi rejeitada todas as vezes pelo experimento, com potência maior do que 90% e valor p maior do que 0.009 para todos os testes, pode-se concluir que, de forma sistemática o estimador *conjunto de alunos de engenharia* produz estimativas ruins.

5. Experimentos Adicionais - Recipiente 2

5.1 As estimativas rejeitam uma hipótese nula mais próxima da real?

5.1.1 Planejamento Experimental

No caso do recipiente 2, não foi possível a coleta de moedas de 5 centavos em quantidade suficiente para permitir a experimentação. Desta forma, foi necessário a realização de aproximações e apenas uma amostra foi possível de ser coletada que será definida com a hipótese nula dos testes para o recipiente 2.

Além disso, para que as estimativas possam ser divididas em grupos, elas deveriam ser independentes, uma vez que se há dependência entre as estimativas, não adiantaria separá-las em grupos para uma conclusão mais generalizada. Para confirmar essa expectativa de dependência entre dados, é proposto o teste Durbin-Watson, que testa a autocorrelação serial [**5**]. Se o resultado para o valor p do teste for significativamente baixo, a hipótese nula de que não há autocorrelação será rejeitada.

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
dwtest(estimacao$Valor~1)
```

```
##
##  Durbin-Watson test
##
## data:  estimacao$Valor ~ 1
```

```
## DW = 1.841, p-value = 0.3324
## alternative hypothesis: true autocorrelation is greater than 0
```

Um valor p de 0.33 é muito alto, indicando que não há evidências suficientes para rejeitar a hipótese de que não há correlação. Desta forma, os grupos poderão ser separados, mas, desta vez, terão que ser testados pela Prova por Contradição, já que não é possível variar o tamanho amostral do experimento.

5.1.1.1 Desrição da coleta de dados

O total de moedas de R$0,05 disponível para a experimentação era de 43. Desta forma, o copo foi enchido

Referências

[1] Patrick Royston (1995) Remark AS R94: A remark on Algorithm AS 181: The W test for normality. Applied Statistics, 44, 547-551

[2] https://www.wyzant.com/resources/lessons/math/prealgebra/representing-data. Acessado em 13/04/2017

[3] D.C. Montgomery, G.C. Runger, Applied Statistics and Probability for Engineers, 5th ed., Wiley, 2010

[4] http://www4.bcb.gov.br/adm/mecir/principal.asp. Acessado em 13/04/2017

[5] Felipe Campelo, Lecture Notes on Design and Analysis of Experiments. UFMG, Belo Horizonte, Brasil (2014)