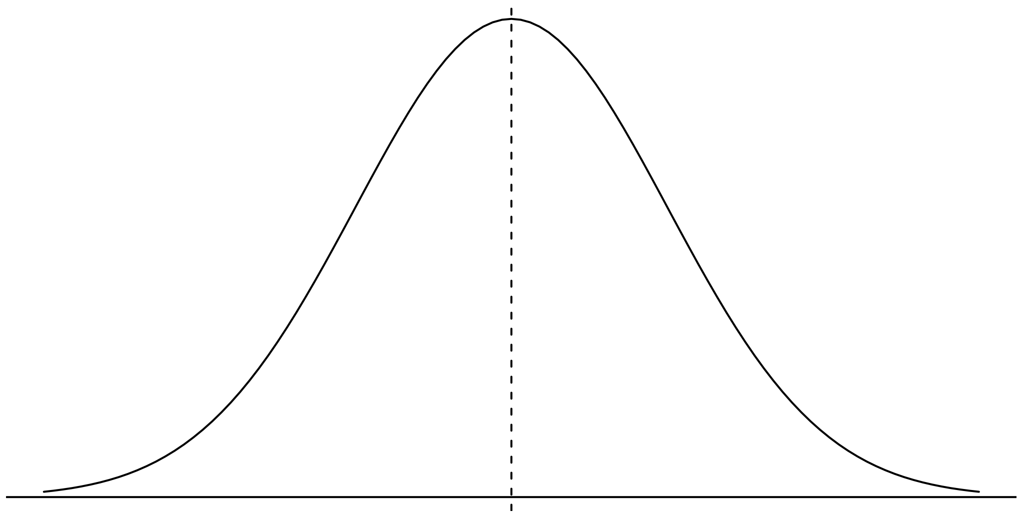
Comparação de testes de normalidade

Anaih P. Pereira e Arthur C. M. Rocha

3 de outubro de 2017



Criando uma função para aplicação dos testes de normalidade:

```
testes<-function(t){

s<-shapiro.test(t)
l<-lillie.test(t)
a<-ad.test(t)
c<-cvm.test(t)
p<-pearson.test(t)
k<-ks.test(t,"pnorm",mean(t),sd(t))

tabela<-data.frame(Teste=c("SW","Lillie","AD","CVM","Chisq","KS"),
                   Estatística=c(s$statistic,l$statistic,a$statistic,
                                 c$statistic,p$statistic,k$statistic),
                   Valor_p=c(s$p.value,l$p.value,a$p.value,
                             c$p.value,p$p.value,k$p.value))
return(tabela)
}
```

Utilizando um banco de dados para aplicar os testes:

Foram feitos os testes para as variáveis peso (com outliers) e peso (sem outliers), idade e altura. A partir dos valores-p obtidos pelos testes, é perceptível que a hipótese nula de normalidade é rejeitada em todos os casos. Os resultados são apresentados nas tabelas abaixo:

- Para os pesos:

```
testes(dados$Peso)   #com outliers
```

```
##    Teste  Estatística      Valor_p
## 1     SW   0.95033571 7.832850e-11
## 2 Lillie   0.08197083 2.644292e-07
## 3     AD   4.40445109 6.019804e-11
## 4    CVM   0.66634124 1.101321e-07
## 5  Chisq 109.59069767 2.332194e-14
## 6     KS   0.08197083 6.186539e-03
```

```
testes(Peso)         #sem outliers
```

```
##    Teste Estatística      Valor_p
## 1     SW  0.96760228 4.655856e-08
## 2 Lillie  0.08601399 6.013929e-08
## 3     AD  3.77567998 1.992169e-09
## 4    CVM  0.58565319 4.773640e-07
## 5  Chisq 83.27186761 1.085340e-09
## 6     KS  0.08601399 3.826119e-03
```

- Para altura:

```
testes(dados$Altura)  #altura
```

```
##    Teste Estatística      Valor_p
## 1     SW  0.95433619 2.854688e-10
## 2 Lillie  0.05966594 8.933214e-04
## 3     AD  1.03273601 1.012611e-02
## 4    CVM  0.13842407 3.377685e-02
```

```
## 5  Chisq 90.33488372 6.477535e-11
## 6     KS  0.05966594 9.361395e-02
```

- Para idade:

```r
testes(dados$Idade)  #idade
```

```
##    Teste Estatística       Valor_p
## 1     SW   0.7861826  2.442673e-23
## 2 Lillie   0.1598575  7.923437e-30
## 3     AD  17.2874648  3.700000e-24
## 4    CVM   2.6076726  7.370000e-10
## 5  Chisq 680.2046512 3.402296e-131
## 6     KS   0.1598575  5.709836e-10
```

Utiizando outro banco de dados para aplicar os testes:

Foram feitos os testes para as variáveis peso, altura, número de filhos. A partir dos valores-p obtidos pelos testes, é perceptível que a hipótese nula de normalidade não é rejeitada para a variável peso. Para a variável altura, o teste qui-quadrado foi o único a rejeitar a hipótese nula, enquanto que para a variável número de filhos todos os testes rejeitaram a hipótese de normalidade. Os resultados são apresentados nas tabelas abaixo:

- Para peso:

```r
testes(dados2$peso)
```

```
##    Teste Estatística   Valor_p
## 1     SW  0.99496046 0.3496246
## 2 Lillie  0.02413216 0.9083980
## 3     AD  0.24366416 0.7633791
## 4    CVM  0.03132502 0.8282354
## 5  Chisq 13.65765766 0.7511198
## 6     KS  0.02413216 0.9901730
```

- Para altura:

```r
testes(dados2$altura)
```

```
##    Teste Estatística      Valor_p
## 1     SW  0.99472258 0.3102498798
## 2 Lillie  0.04745225 0.0682086686
## 3     AD  0.54392554 0.1611879381
## 4    CVM  0.09594664 0.1273451103
## 5  Chisq 42.79279279 0.0008559786
## 6     KS  0.04745225 0.4414589884
```

- Para numero de filhos:

```r
testes(dados2$Numerofilhos)
```

```
##    Teste Estatística      Valor_p
## 1     SW   0.8710386 1.991316e-08
## 2 Lillie   0.2292453 6.465507e-16
## 3     AD   6.8839748 5.814830e-17
## 4    CVM   1.2440423 7.370000e-10
## 5  Chisq 412.7500000 1.199785e-81
## 6     KS          NA           NA
```

Simulações para achar o poder dos testes:

- Definições Gerais

```
set.seed(12345)

namostra<-round(seq(10,2000,,25))  #tamanhos da amostra

poder<-list() #lista vazia para guardar os poderes
```

- A partir de simulações, foram geradas amostras de tamanhos entre 0 e 2000 de uma v.a Beta(2,2) e aplicados os testes de normalidade, a fim de obter-se o poder do teste (1-erro do tipo II), para isso extraiu-se os valores-p dos testes e contabilizaram-se a quantidade de testes onde a hipótese nula (normalidade) não foi rejeitada e assim, fazendo a diferença de 1 e essa quantidade, estimou-se empiricamente o poder dos testes.

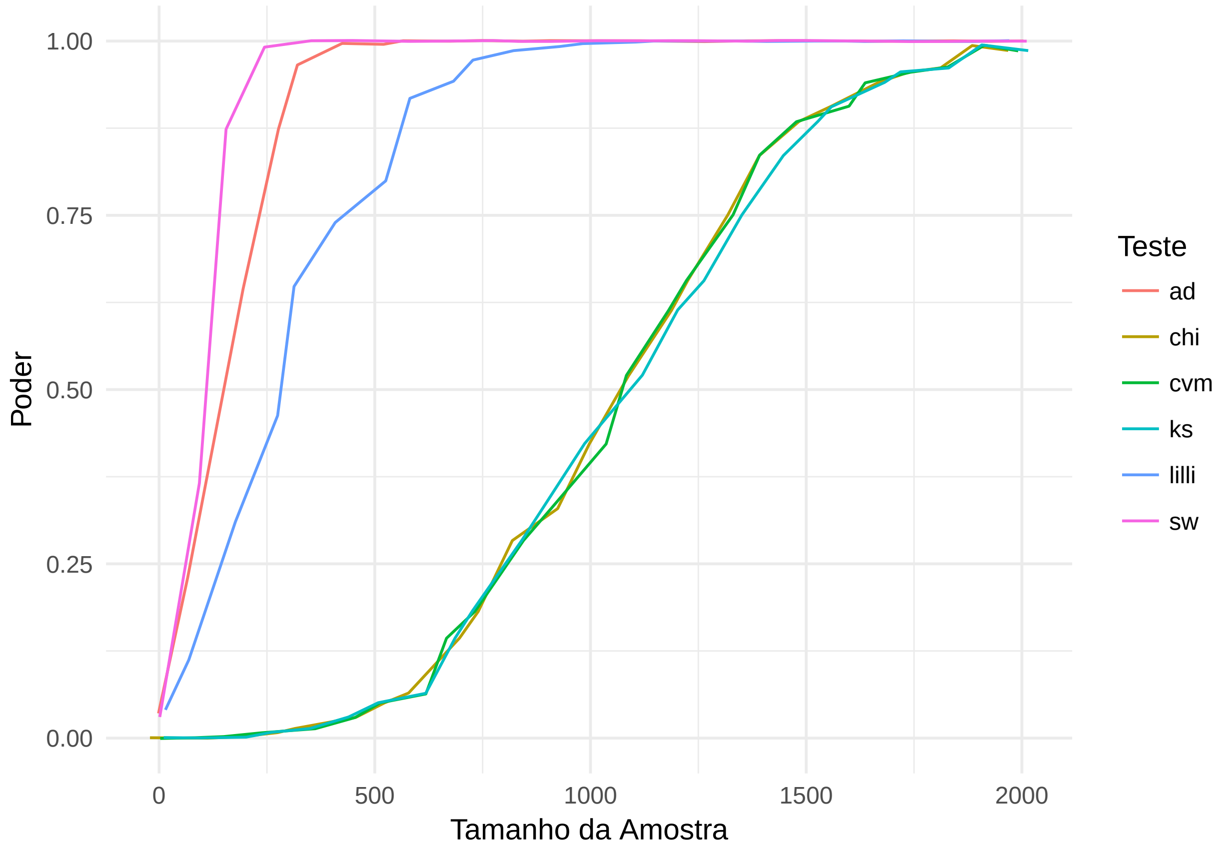
- Os códigos a seguir representam as simulações:

```
for(i in 1:length(namostra)){

  boot<-replicate(500,{
    x<-rbeta(namostra[i],2,2)
    ks<-ks.test(x,"pnorm",mean(x),sd(x))$p.value
    sw<-shapiro.test(x)$p.value
    lilli<-lillie.test(x)$p.value
    ad<-ad.test(x)$p.value
    cvm<-cvm.test(x)$p.value
    chi<-pearson.test(x)$p.value
    c(ks,sw,lilli,ad,cvm,chi)
  })


  b<-apply(boot,1,function(x)return(x>.05))
  testes<-apply(b,2,mean)
  names(testes)<-c('ks','sw','lilli','ad',"cvm","chi")
poder[[i]]<-1-testes
}
```

Gráfico dos poderes

Usando o pacote ggplot2, fez-se o gráfico dos poderes dos testes conforme o número de amostra

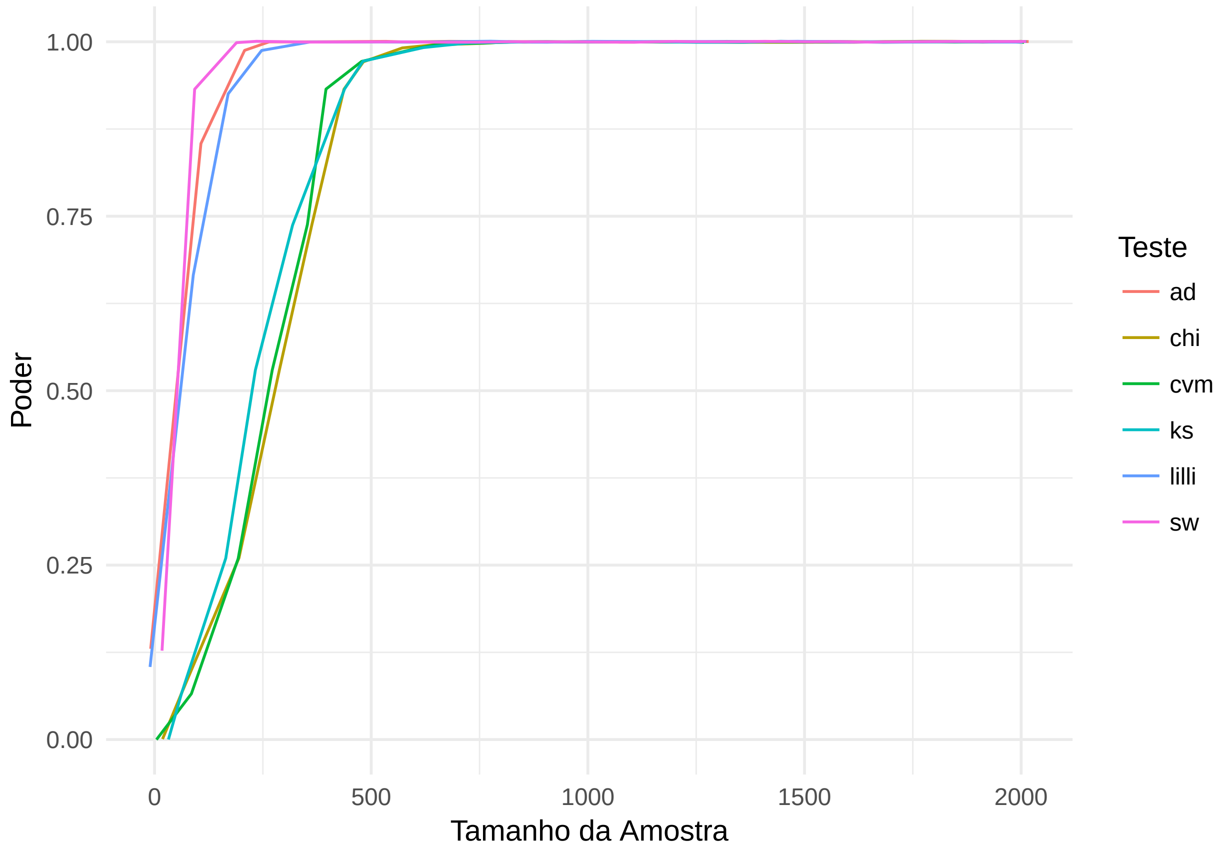
```
library(ggplot2)

#criando o data frame para o ggplot
dt<-data.frame(rep(namostra,6),c(ks,sw,lilli,ad,cvm,chi))
dt$teste<-rep(c("ks","sw","lilli","ad","cvm","chi"),each=25)
names(dt)<-c("namostra","Poder","Teste")
plot<-ggplot(dt,aes(namostra,Poder,group=Teste,col=Teste))+geom_line(position="jitter")
plot<-plot+theme_minimal() + xlab("Tamanho da Amostra")
plot
```



É perceptível que a classificação quanto ao poder foi:

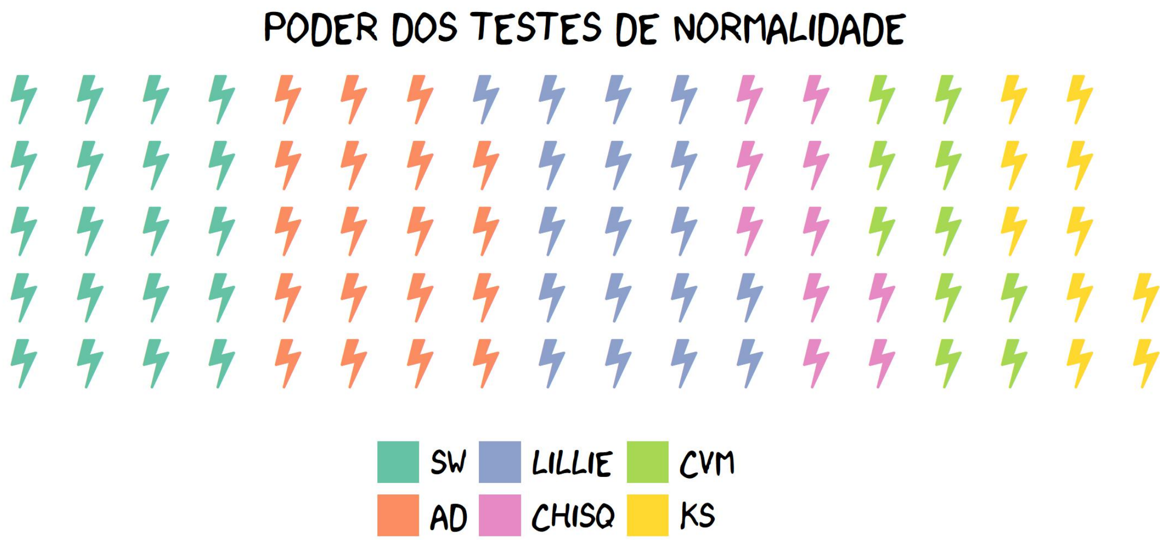
```
## [1] "sw"    "ad"    "lilli" "chi"   "cvm"   "ks"
```

- Análogamente, fez-se simulações de amostras de uma v.a com distribuição Gamma(4,5), obtendo-se o seguinte resultado:



Resultados:

- Observou-se, pelos estudos que a classificação de poder é:



*

*