Questoes.R

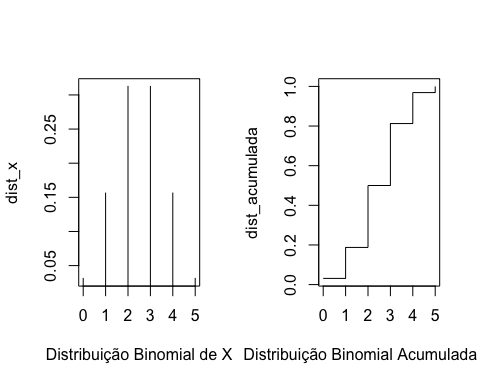
edevaldogaudencio

2020-12-09

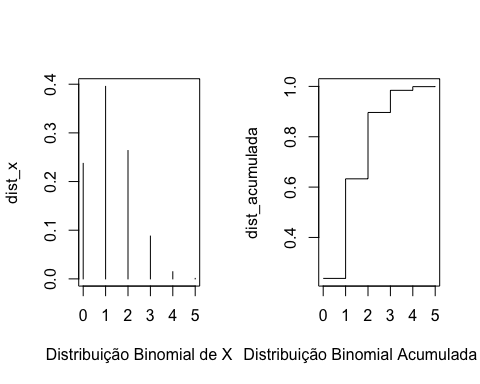
####### Lista 2 - Estatística  
####### Mestrado Profissionalizante em Economia  
####### Aluno: Edevaldo Siqueira Gaudencio  
  
####### Carregando pacotes  
  
  
########Questão 1  
# Um fabricante de peças de automóveis garante que uma caixa de suas peças conterá no   
# máximo, duas defeituosas. Se a caixa contém 18 peças, e sabendo que a experiência mostra   
# que 5% das peças produzidas apresentam defeito, qual a probabilidade de que uma caixa   
# satisfaça a garantia?   
  
# Dica use a função dbinom ou no excel DISTR.BINOM(número de sucesso;  
# número de tentativas; probabilidade; cumulativo=1) para calcular a probabilidade de 0 peças,   
# 1 peça ou 2 peças defeituosas e calcule a função de distribuição acumulada até 2.)  
  
 # Interpretando o enunciado:  
 # Amostra=18  
 # Defeito=5% (0.05)  
 # Tentativas=2  
 prob\_0 <- dbinom(0, 18, 0.05)  
 prob\_1 <- dbinom(1, 18, 0.05)  
 prob\_2 <- dbinom(2, 18, 0.05)  
 sum(prob\_0+prob\_1+prob\_2)

## [1] 0.9418711

########Questão 2  
# Se X tem distribuição binomial com parâmetros n=5 e p=½, faça os gráficos da   
# distribuição de X e a função de distribuição acumulada F(X).   
  
 # Interpretando o enunciado:  
 # n=5  
 # p=0.5  
 par(mfrow = c(1, 2))  
 x <- 0:5  
 dist\_x <- dbinom(x, size = 5, prob = 0.5)  
 plot(x, dist\_x, type = "h", xlab = "Distribuição Binomial de X")  
 dist\_acumulada <- pbinom(x, size = 5, prob = 0.5)  
 plot(x, dist\_acumulada, type = "s", xlab = "Distribuição Binomial Acumulada de X")



par(mfrow = c(1, 2))  
  
  
## item A:  
# Considere n=5 e p=¼. Obtenha o gráfico da distribuição de X.   
# Qual a diferença do resultado do exercício anterior?  
  
 # Interpretando o enunciado:  
 # n=5  
 # p=0.25  
 par(mfrow = c(1, 2))  
 x <- 0:5  
 dist\_x <- dbinom(x, size = 5, prob = 0.25)  
 plot(x, dist\_x, type = "h", xlab = "Distribuição Binomial de X")  
 dist\_acumulada <- pbinom(x, size = 5, prob = 0.25)  
 plot(x, dist\_acumulada, type = "s", xlab = "Distribuição Binomial Acumulada de X")



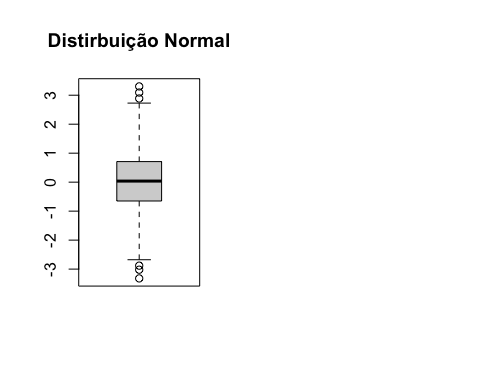
par(mfrow = c(1, 2))  
  
  
########Questão 3 - Distribuição Normal   
# xi é uma variável aleatória com distribuição normal padrão (média zero e variância 1).  
## Item A  
# Qual a média e variância de uma distribuição Normal (0,1)? Gere mil observações   
# dessa distribuição Normal (0,1). Calcule a média, variância, 1o. Quartil, media,   
# 3o. Quartil dessas observações geradas. Faça um box-plot e histograma dos dados   
# gerados, qual a sua conclusão com a comparação distribuição teórica.   
# (Dica: No excel vá em Dados/Análise de Dados/Gerar Números Aleatórios.   
# Selecione 1000 números e distribuição normal. No r, use a função rnorm)   
  
 # Interpretando o enunciado:  
 #media=0  
 #variancia=1  
 #amostra=1000  
   
 # Gerando amostra de mil observações  
 set.seed(100)  
 dist\_normal <- rnorm(1000,0,1)  
   
 # Média, 1o. Mediana, Quartil, 3o. Quartil  
 summary(dist\_normal)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -3.32078 -0.64970 0.03690 0.01681 0.70959 3.30415

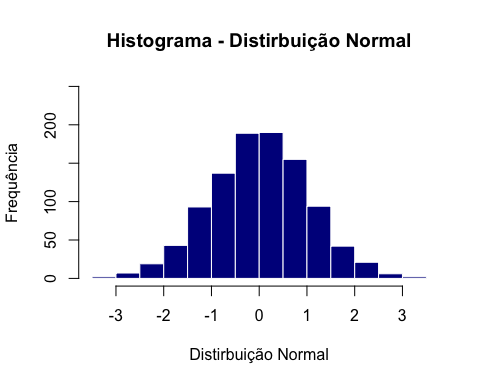
# Variância  
 var(dist\_normal)

## [1] 1.062112

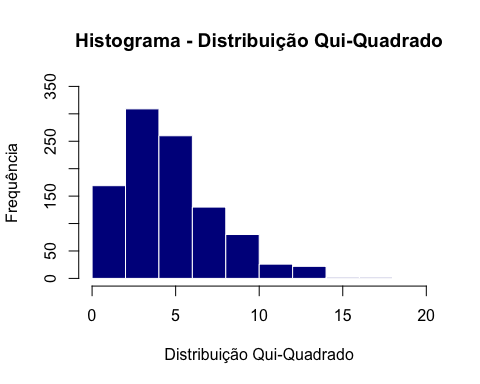
# Faça um box-plot dos dados gerado.  
 boxplot(dist\_normal, main = "Distirbuição Normal")  
   
 # Faça um histograma dos dados gerado.  
 par(mfrow = c(1, 1))



hist(dist\_normal, col = "darkblue", xlab = "Distirbuição Normal",   
 ylab = "Frequência", border = "white", main = "Histograma - Distirbuição Normal",  
 ylim=c(0, 250))



## Item B  
# Gere mais quarto distribuições de uma distribuição normal padrão com 1000 observações.   
# Assim, ficaremos com 5 distribuições normais padrões independentes. (Se manteve a   
# semente fixa, lembre-se de alterá-la). Eleve ao quadrado cada uma das séries e some   
# os valores das 5 séries geradas. Essa soma, de acordo com os resultados acima, deve   
# ter distribuição Y ~ 2(5), ou seja, distribuição chi quadrado com 5 graus de   
# liberdade. Sabemos também que E(Y)=número de graus=5 e Var(Y)=2\*número de graus   
# de liberdade=10, uma vez que número de graus de liberdade de X é 5. Faça um   
# histograma, calcule a média e variância. Compare os resultados com os valores   
# teóricos da distribuição 2(5).   
  
 # Interpretando o enunciado:  
 # 5 distribuições normais independentes  
 # Elevar ao quadrado   
 # Some os valores das 5 séries geradas  
 # Faça um histograma  
 # Calcule a média e variância.  
   
   
 # Gerando 5 amostras de mil observações cada  
 set.seed(101)  
 dist\_normal\_1 <- rnorm(1000,0,1)   
 set.seed(102)  
 dist\_normal\_2 <- rnorm(1000,0,1)   
 set.seed(103)  
 dist\_normal\_3 <- rnorm(1000,0,1)   
 set.seed(104)  
 dist\_normal\_4 <- rnorm(1000,0,1)   
 set.seed(105)  
 dist\_normal\_5 <- rnorm(1000,0,1)   
   
 # Elevar ao quadrado cada uma das distribuições e somar os valores das 5 séries geradas  
 qui\_quadrado <- (dist\_normal\_1^2+dist\_normal\_2^2+dist\_normal\_3^2+dist\_normal\_4^2+dist\_normal\_5^2)   
   
 # Faça um histograma  
 par(mfrow = c(1, 1))  
 hist(qui\_quadrado , col = "darkblue", xlab = "Distribuição Qui-Quadrado",   
 ylab = "Frequência", border = "white", main = "Histograma - Distribuição Qui-Quadrado",  
 ylim=c(0, 350), xlim=c(0, 20))



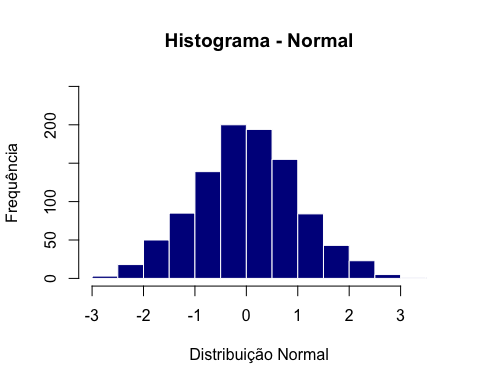
# Média, 1o. Mediana, Quartil, 3o. Quartil  
 summary(qui\_quadrado)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.09483 2.50597 4.10946 4.63635 6.11548 17.23165

# Variância  
 var(qui\_quadrado)

## [1] 8.231508

# Comparando com os valores teóricos, percebe-se que a variancia tende a 10 e a média tende a 5.   
 # Realizada uma simulação com 10.000 números em cada distribuição normal, os valores foram ainda  
 # mais próximos do referencial teórico: média de 4.94 e variancia de 9.67  
   
  
## Item C  
# Gere agora uma normal X~ N(5,10), primeiro vamos padronizar essa série, isto é, vamos gerar   
# uma nova série da seguinte forma:   
# Calcule a média e desvio-padrão de z, faça um histograma de z.   
# De fato, podemos dizer que Z~N(0,1)\_.  
   
 # Interpretando o enunciado:   
 # Média: 5  
 # Variancia: 10  
 # Amostra: 1000  
   
 # Gerando amostra de mil observações  
 set.seed(106)  
 dist\_normal\_6 <- rnorm(1000,5,sqrt(10))  
   
 #vamos padronizar essa série: Z=(X-μ)/√(σ^2 )  
 dist\_normal\_6\_padronizada <- (dist\_normal\_6 - 5)/sqrt(10)  
   
 # Faça um histograma  
 par(mfrow = c(1, 1))  
 hist(dist\_normal\_6\_padronizada, col = "darkblue", xlab = "Distribuição Normal",   
 ylab = "Frequência", border = "white", main = "Histograma - Normal",  
 ylim=c(0, 250))



# Média, 1o. Mediana, Quartil, 3o. Quartil  
 summary(dist\_normal\_6\_padronizada)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.88114 -0.64500 0.01537 0.01381 0.65205 3.03446

# Desvio padrão  
 sd(dist\_normal\_6\_padronizada)

## [1] 0.9930022

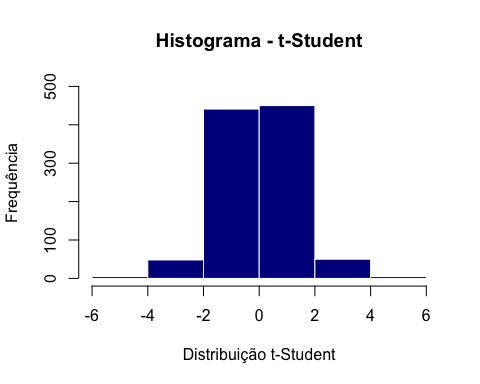
# A média tende a zero e o desvio padrão tende a 1.  
 # Realizada uma simulação com 100.000 números na distribuicao , os valores foram ainda  
 # mais próximos do referencial teórico: média de -0.0004 e devio padrão de 0.9963.  
   
   
## Item D  
# Por fim, vamos simular uma distribuição t(5) com 5 graus de liberdade.   
# Use as séries geradas no item para calcular:t(5)=Z/√(Y/5),  
# Use os valores gerados no item c) para o numerador e os valores gerados no item b)   
# para o denominador. Calcule média e variância dessa nova distribuição.   
# Faça um histograma e um box-plot. Há diferenças para a normal (0,1)?   
   
 # Interpretando o enunciado:   
 # simular uma distribuição t(5) com 5 graus de liberdade  
 # Valores do item C para o numerador e B para denominador  
 # Calcule média e variância dessa nova distribuição  
 # Faça um histograma e um box-plot  
 # Diferenças em relação a distribuição normal (0,1?  
   
 # simular uma distribuição t(5)  
 dist\_t\_student <- (dist\_normal\_6\_padronizada/(sqrt(qui\_quadrado/5)))  
   
 # Média, 1o. Mediana, Quartil, 3o. Quartil  
 summary(dist\_t\_student)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -18.280775 -0.770148 0.023386 0.003465 0.707233 5.309786

# Variância  
 var(dist\_t\_student)

## [1] 1.970871

# Faça um histograma  
 par(mfrow = c(1, 1))  
 hist(dist\_t\_student, col = "darkblue", xlab = "Distribuição t-Student",   
 ylab = "Frequência", border = "white", main = "Histograma - t-Student",  
 xlim=c(-6, 6), ylim=c(0, 500))



# A média tende a zero e a variancia tende a 2.  
 # Na distribuição normal, a variância tendia a 1  
  
  
########Questão 4 - Distribuição Normal   
# Explique sucintamente as seguintes definições de estimadores abaixo:  
  
 # A) estimador não-viesado: Um estimador T é não-viesado (ou não   
 # tendencioso) se seu valor esperado for o próprio parâmetro Q que   
 # se pretende estimar, isto é, E(T) = Q   
   
 # B) estimador consistente: Consistência é uma propriedade por meio da   
 # qual a acurácia deuma estimativa aumenta quando o tamanho da amostra   
 # aumenta. Dizemos que um estimador para o parâmetro é consistente se,   
 # além de ser não-viesado, sua variância tende a zero quando o tamanho  
 # amostral tende a infinito.  
   
 # C) Explique quando um estimador A é mais eficiente que um estimador B:   
 # Caso o estimador A e B sejam não viesados de um mesmo parâmetro Q,   
 # o primeiro será mais efienciente do que o segundo quando a variância   
 # de A for menor que a variância de B.  
  
########Questão 5 - Descreva os seguintes conceitos e dê exemplo em cada um dos   
 # itens:  
 # A) Erro tipo 1 e tipo 2 de um teste: Nenhum teste de hipótese é 100% certo.   
 # O teste é baseado em probabilidades e sempre há uma possibilidade de   
 # chegar em uma conclusão errada. Os dois erros possíveis são: tipo 1   
 # e tipo 2. Os dois tipos são inversamente relacionados e são   
 # determinados pelo nível de significancia e o poder do teste.   
 # Logo, é necessário definir qual erro tem consequências mais severas   
 # para a situação antes de definir os riscos.   
 #  
 # Tipo 1: quando a hipótese nula é verdadeira e você a rejeita,   
 # comete um erro do tipo I. A probabilidade de cometer esse tipo de erro  
 # é o nível de significância que foi definido para o teste de hipóteses.  
 # Para reduzir o risco de cometer esse erro, você precisa utilizar um   
 # valor inferior para o nível de significancia. Exemplo:  
 #  
 # Tipo 2: Quando a hipótese nula é falsa e você não a rejeita,   
 # comete um erro de tipo II. A probabilidade de cometer esse erro depende  
 # do poder do teste. Para reduzir o risco de cometer esse erro,   
 # é necessário garantir que o teste tenha potência suficiente, em outras  
 # palavras,o tamanho amostral seja grande o suficiente para detectar   
 # uma diferença prática, quando realmente existir uma.  
   
 # B) poder de um teste de hipótese: O Poder do Teste tem como objetivo   
 # conhecer o quanto o teste de hipóteses controla um erro do tipo II,   
 # ou qual a probabilidade de rejeitar a hipótese nula se realmente   
 # for falsa.  
 # C) p-valor: O valor-p indica a probabilidade de se observar uma diferença   
 # tão grande ou maior do que a que foi observada sob a hipótese nula.  
   
  
   
########Questão 6) Explique de forma breve e intuitiva as diferentes formas   
 # de estimação:  
 # i) Estimador de Momentos (mostre o estimador de momento para a média   
 # e variância): O método dos momentos consiste em igualar os momentos   
 # amostrais aos populacionais. O resultado dessa operação produzirá   
 # as estimativas dos parâmetros da distribuição de probabilidades   
 # em questão. Seja X uma variável aleatória com média μ e   
 # variância σ^2. Neste caso, as seguintes relaçõe são válidas   
 # para os dois primeiros momentos populacionais:  
 # E(X)= μ, E=(X^2 )=σ^2+ μ^2, do qual obtemos:   
 # μ=E(X), σ^2=(X^2 )-E^2 (X)   
 #  
   
   
 # ii) Estimador de mínimos quadrados: é uma técnica de otimização   
 # matemática que procura encontrar o melhor ajuste para um conjunto   
 # de dados tentando minimizar a soma dos quadrados das diferenças entre   
 # o valor estimado e os dados observados.É a forma de estimação mais   
 # amplamente utilizada na econometria.   
   
   
 # iii) Estimador de máxima verossimilhança: método para estimar os   
 # parâmetros de um modelo estatístico. Assim, a partir de um conjunto   
 # de dados e, dado um modelo estatístico, a estimativa por máxima   
 # verossimilhança estima valores para os diferentes parâmetros do   
 # modelo, buscando maximizar a probabilidade dos dados observados.   
 # Apresenta-se como um método geral para estimação de parâmetros,   
 # principalmente no caso de distribuições normais.   
   
########Questão 8)   
# Uma v.a X tem distribuição normal com média 10 e desvio-padrão 4. Imagine   
# que um jogo premie toda amostra cuja média é maior 12.   
  
## Item A   
# Se um participante escolher uma amostra de tamanho 16, qual é a probabilidade   
# de ele ganhar o prêmio?  
 media <- 10  
 desvio <-4  
 n <- 16  
 erro\_padrao <- desvio/sqrt(n)  
 1-pnorm(12,media,erro\_padrao)

## [1] 0.02275013

## Item B   
# Escolha um tamanho de amostra menor que 16 para participar do jogo. Qual a   
# probabilidade de você ganhar prêmio??  
 # Amostra de 15 observações  
 media <- 10  
 desvio <-4  
 n <- 1:15  
 erro\_padrao <- desvio/sqrt(n)  
 1-pnorm(12,media,erro\_padrao)

## [1] 0.30853754 0.23975006 0.19323812 0.15865525 0.13177624 0.11033568  
## [7] 0.09293837 0.07864960 0.06680720 0.05692315 0.04862721 0.04163226  
## [13] 0.03571173 0.03068441 0.02640376

## Item C  
# Baseado nos resultados qual o melhor tamanho de amostra para participar   
# do jogo?   
 # Conforme pode ser observado, a amostra de tamanho 1 oferece a maior  
 # de o jogador ganhar: 30,85%.  
   
  
########Questão 9)   
# Um professor aplica um teste rápido para seus alunos de 20 questões do tipo   
# certo-errado. O professor coloca como critério de aprovação a seguinte regra   
# “Para ser aprovado o aluno precisa acertar ao menos 13 questões”. Qual é a   
# probabilidade de o aluno ser aprovado, apenas marcando as questões ao acaso?   
# Imagine agora que o professor queira que essa probabilidade de ser aprovado   
# marcando questões ao acaso seja menor que 5%, como ele deveria alterar a   
# regra de aprovação?   
  
 # Interpretando o enunciado:  
 # Questoes=20  
 # Probabilidade=0,5  
 # Sucessos=13   
   
 # Qual é a probabilidade de o aluno ser aprovado, apenas marcando as   
 # questões ao acaso?  
 # probabilidade de >13 é igual 1 - probabilidade acumulada de <12:  
 1-pbinom(12, 20, 0.5)

## [1] 0.131588

# Aprovação < que 5%  
 1-pbinom(13, 20, 0.5) # >=14

## [1] 0.05765915

1-pbinom(14, 20, 0.5) # >=15

## [1] 0.02069473

1-pbinom(15, 20, 0.5) # >=16

## [1] 0.005908966

1-pbinom(16, 20, 0.5) # >=17

## [1] 0.001288414

# Como pode ser observado, para reduzir a probabilidade de aprovação com  
 # marcação ao acaso para <5%, o professor deve alterar a regra de aprovação   
 # para 15 acertos ou mais.