

Aluno: Marcelo Melo de Oliveira

Link do Repositório no Github: <https://github.com/marcelomeloo/P1-GED13/tree/master>

Questão 1

a)

- Inicializacao

```
library(nortest)
amostra <- c(149.3355, 140.3779, 145.7254, 149.8931, 139.6168, 149.1934,
129.6147, 134.7523, 167.8030, 171.7407, 157.5422, 160.2664, 155.4553,
142.5989, 134.9844, 148.5172, 163.1447, 131.0138, 130.2423, 167.2239,
149.4015, 145.6802, 160.3472, 121.1775, 136.7295, 162.2381, 150.7192,
117.8144, 137.3630, 158.6373, 168.0833, 133.9263, 150.9102, 149.4811,
167.4367, 178.0970, 138.4903, 148.6764, 181.0990, 167.3345, 147.0679,
156.1410, 148.8734, 140.9484, 147.6408, 134.5726, 184.6812, 134.6648,
146.8130, 167.4161)

z.amostra <- scale(amostra)
```

- Testes

```
ks.test(z.amostra, "pnorm", 0, 1) #p-value = 0.4688

##
## Exact one-sample Kolmogorov-Smirnov test
##
## data: z.amostra
## D = 0.1167, p-value = 0.4688
## alternative hypothesis: two-sided

shapiro.test(amostra) #p-value = 0.6324

##
## Shapiro-Wilk normality test
##
## data: amostra
## W = 0.98185, p-value = 0.6324

ad.test(amostra) #p-value = 0.3928

##
## Anderson-Darling normality test
##
## data: amostra
## A = 0.37902, p-value = 0.3928
```

```

lillie.test(amostra) #p-value = 0.08619

##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data: amostra
## D = 0.1167, p-value = 0.08619

```

- Interpretacao

Como em todos os testes obteve-se um valor de p maior do que 0.05, nao foi negada a hipotese nula dos testes de normalidade. Isto eh, eh possivel que a amostra siga uma distribuicao normal.

b)

```

mi <- 150
sigma <- 15

phi <- function(x) (1/(sigma*sqrt(2*pi)))*exp((-1/2)*((x-mi)/sigma)^2)
plot(phi, 100, 200, lwd=5, xlab="segundos", ylab="freq")
x <- seq(125, 150, 0.1)
lines(x, (1/(sigma*sqrt(2*pi)))*exp((-1/2)*((x-mi)/sigma)^2), type="h", col="grey")

```

Logo a probabilidade desejada deve ser:

```

pnorm(150,mean=mi,sd=sigma)-pnorm(125,mean=mi,sd=sigma)
## [1] 0.4522096

```

c)

```

phi <- function(x) (1/(sigma*sqrt(2*pi)))*exp((-1/2)*((x-mi)/sigma)^2)
plot(phi, 100, 200, lwd=5, xlab="segundos", ylab="freq")
x <- seq(100, 125, 0.1)
lines(x, (1/(sigma*sqrt(2*pi)))*exp((-1/2)*((x-mi)/sigma)^2), type="h", col="grey")

```

Logo a probabilidade desejada deve ser:

```

pnorm(125,mean=mi,sd=sigma)-pnorm(100,mean=mi,sd=sigma)
## [1] 0.04736129

```

d)

```

phi <- function(x) (1/(sigma*sqrt(2*pi)))*exp((-1/2)*((x-mi)/sigma)^2)
plot(phi, 100, 200, lwd=5, xlab="segundos", ylab="freq")
x <- seq(145, 155, 0.1)
lines(x, (1/(sigma*sqrt(2*pi)))*exp((-1/2)*((x-mi)/sigma)^2), type="h", col="grey")

```

Logo a probabilidade desejada deve ser:

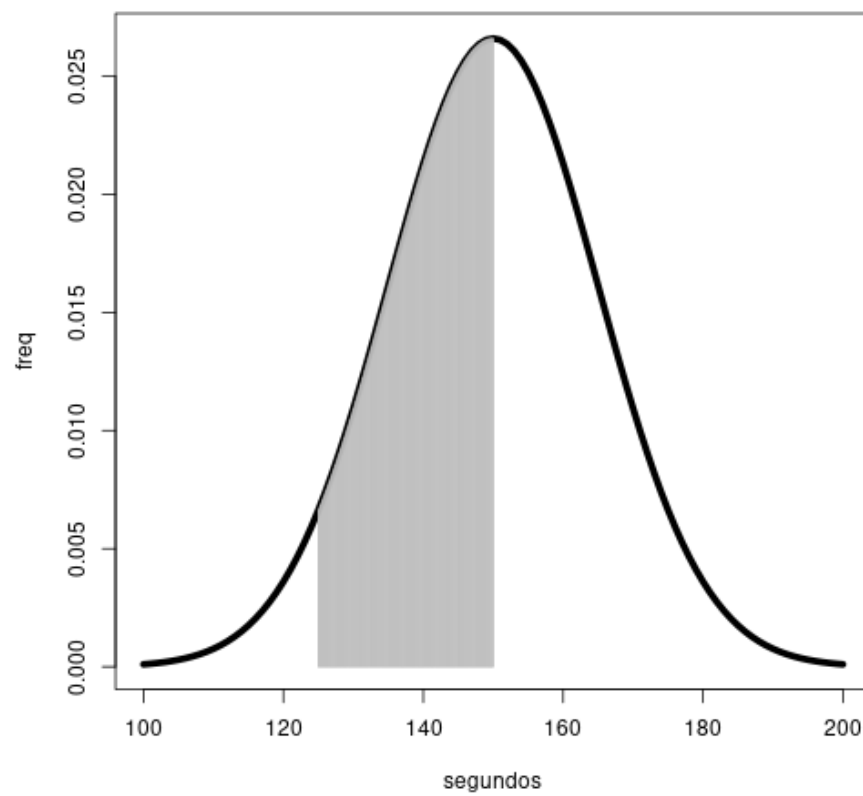


Figure 1: plot of chunk Probabilidade de que uma chamada demore entre 125 e 150 segundos.

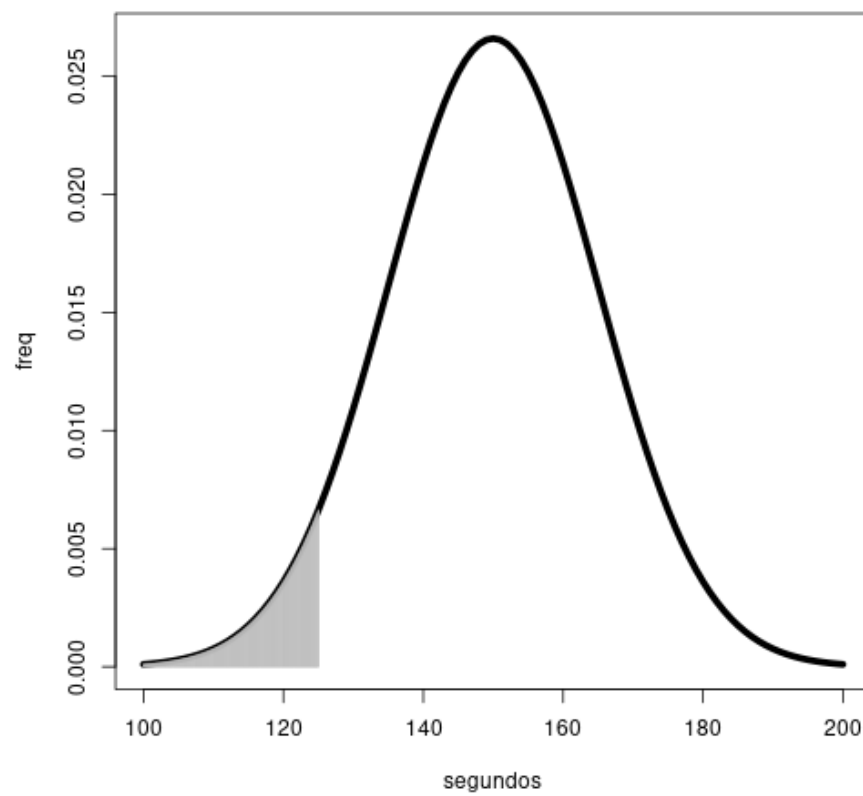


Figure 2: plot of chunk Probabilidade de que uma chamada demore menos de 125 segundos.

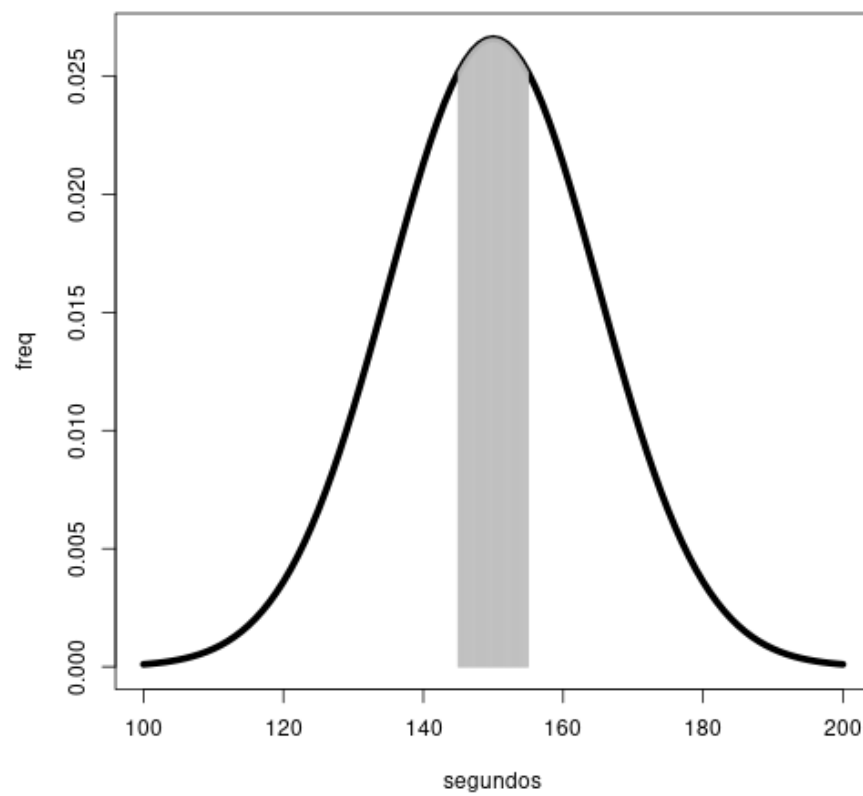


Figure 3: plot of chunk Probabilidade de que uma chamada demore entre 145 e 155 segundos.

```
pnorm(155,mean=mi,sd=sigma)-pnorm(145,mean=mi,sd=sigma)
## [1] 0.2611173
```

e)

```
phi <- function(x) (1/(sigma*sqrt(2*pi)))*exp((-1/2)*((x-mi)/sigma)^2)
plot(phi, 100, 200, lwd=5, xlab="segundos", ylab="freq")
x <- seq(160, 165, 0.1)
lines(x, (1/(sigma*sqrt(2*pi)))*exp((-1/2)*((x-mi)/sigma)^2), type="h", col="grey")
```

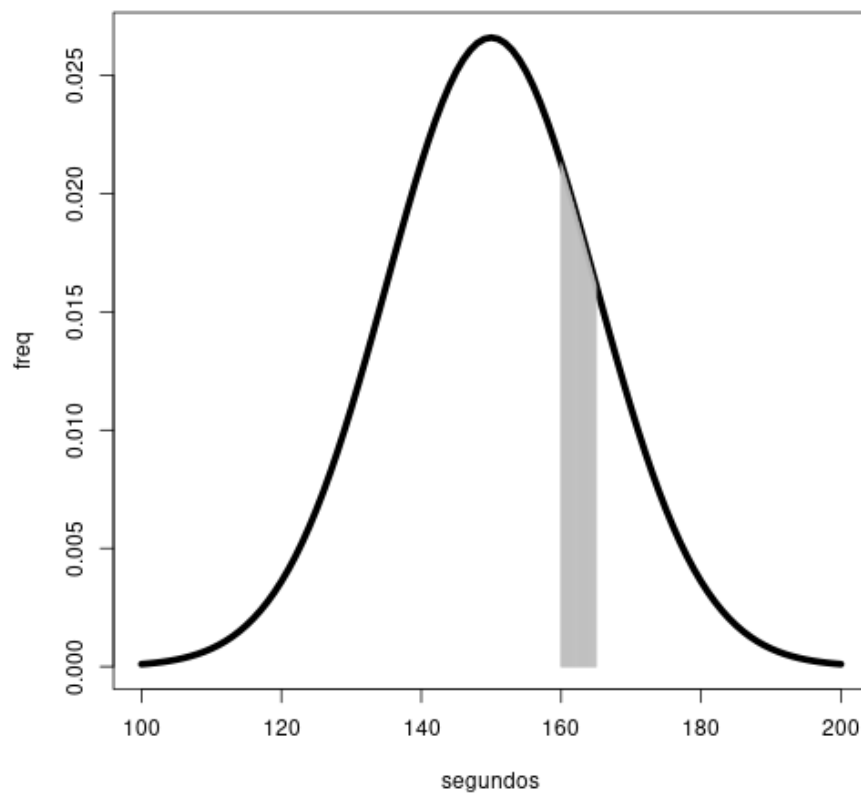


Figure 4: plot of chunk Probabilidade de que uma chamada demore entre 160 e 165 segundos.

Logo a probabilidade desejada sera:

```
pnorm(165,mean=mi,sd=sigma)-pnorm(160,mean=mi,sd=sigma)
```

```
## [1] 0.09383728
```

Questão 2

a) Inicializacao

```
library(fitdistrplus)

## Loading required package: MASS
## Loading required package: survival

library(logspline)

dados_1 <- c(1.5041523,0.5808705,0.6199756,0.2929858,1.2147418,1.1732594,1.8467931,0.72696
dados_2 <- c(0.135745877,0.027801227,0.621652353,0.095321044,1.869870340,0.014069777,0.504
```

b)

Utilizando o comando descdist e mostre as estatísticas geradas.

```
descdist(dados_1, discrete = FALSE, graph = FALSE)

## summary statistics
## -----
## min:  0.1266975   max:  5.296872
## median:  1.307606
## mean:  1.48898
## estimated sd:  0.8704146
## estimated skewness:  1.159633
## estimated kurtosis:  4.644712

descdist(dados_2, discrete = FALSE, graph = FALSE)

## summary statistics
## -----
## min:  0.001452129   max:  1.86987
## median:  0.20739
## mean:  0.3038428
## estimated sd:  0.3081601
## estimated skewness:  2.057199
## estimated kurtosis:  8.673908
```

c)

Utilizando o comando fitdist para o primeiro conjunto de dados, temos:

```
fit.weibull_1 <- fitdist(dados_1, "weibull")
fit.gamma_1 <- fitdist(dados_1, "gamma")
fit.lognormal_1 <- fitdist(dados_1, "lnorm")
fit.normal_1 <- fitdist(dados_1, "norm")
```

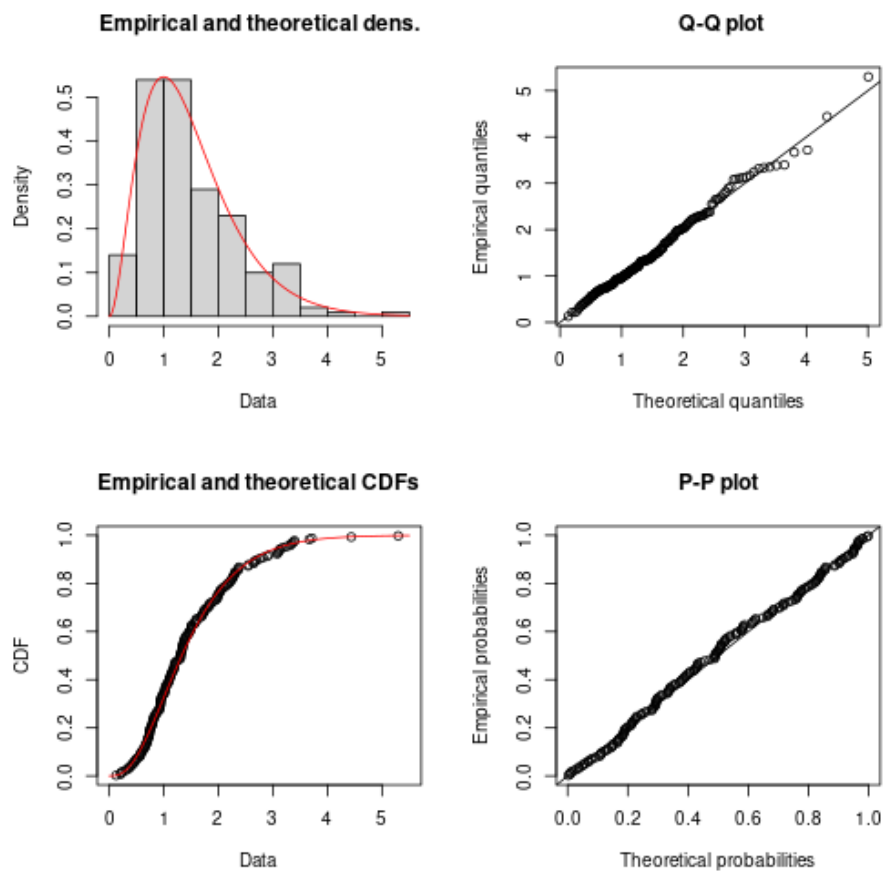
```

fit.expo_1 <- fitdist(dados_1, "exp")
fit.gamma_1

## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters:
##      estimate Std. Error
## shape 3.027141  0.2875985
## rate  2.032940  0.2100724

plot(fit.gamma_1)

```



Eh provavel que o primeiro conjunto de dados seja melhor descrito pela distribuicao gama, uma vez que eh a distribuicao que apresenta menor erro relativo para o ajuste e possui melhor alinhamento as retas dos graficos

Agora, utilizando o comando fitdist para o segundo conjunto de dados, temos:

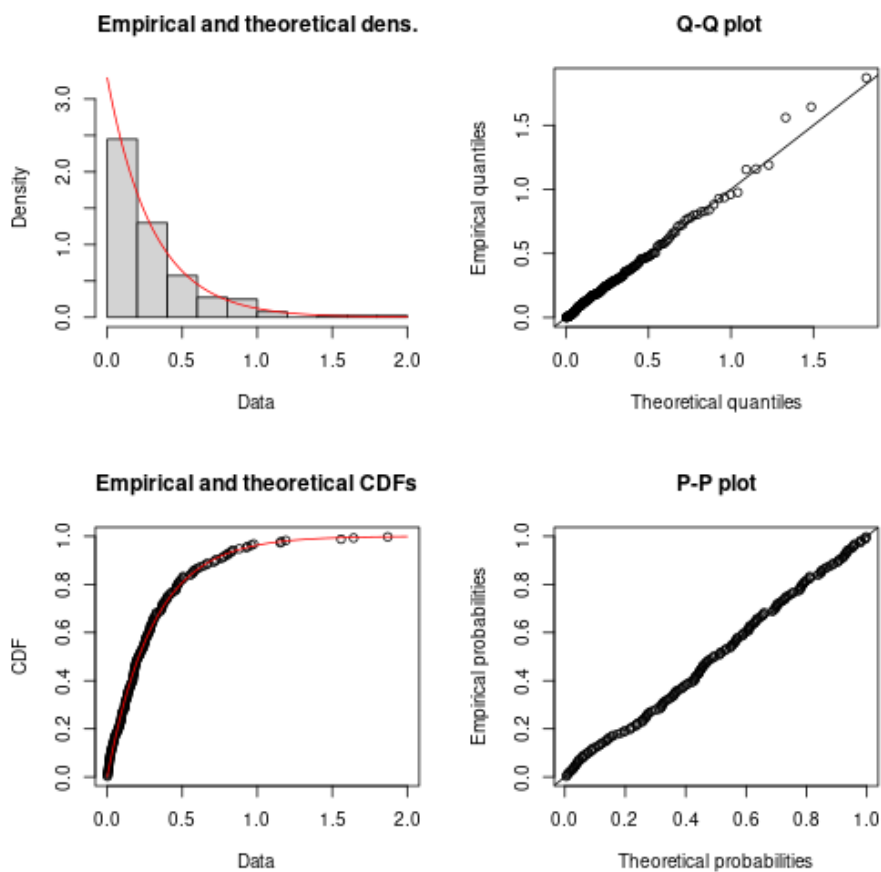

```

fit.weibull_2 <- fitdist(dados_2, "weibull")
fit.gamma_2 <- fitdist(dados_2, "gamma")
fit.lognormal_2 <- fitdist(dados_2, "lnorm")
fit.normal_2 <- fitdist(dados_2, "norm")
fit.expo_2 <- fitdist(dados_2, "exp")
fit.expo_2

## Fitting of the distribution ' exp ' by maximum likelihood
## Parameters:
##      estimate Std. Error
## rate 3.291175  0.2327212

plot(fit.expo_2)

```

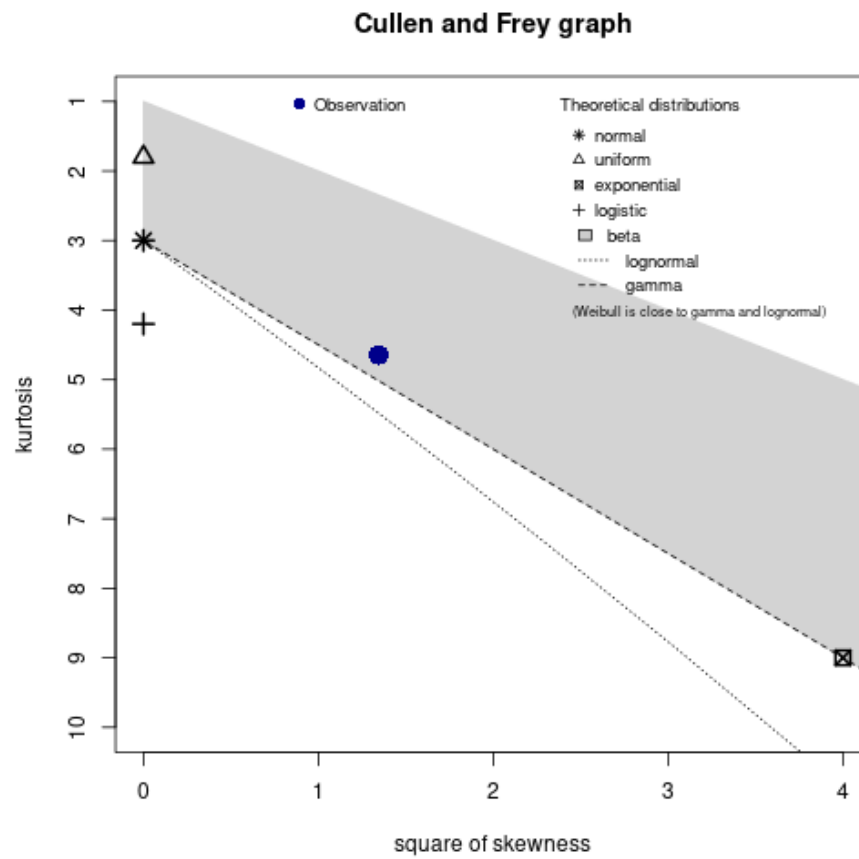


Eh provavel que o segundo conjunto de dados seja melhor descrito pela distribuicao exponencial, uma vez que eh a distribuicao possui melhor alinhamento as retas dos graficos

d)

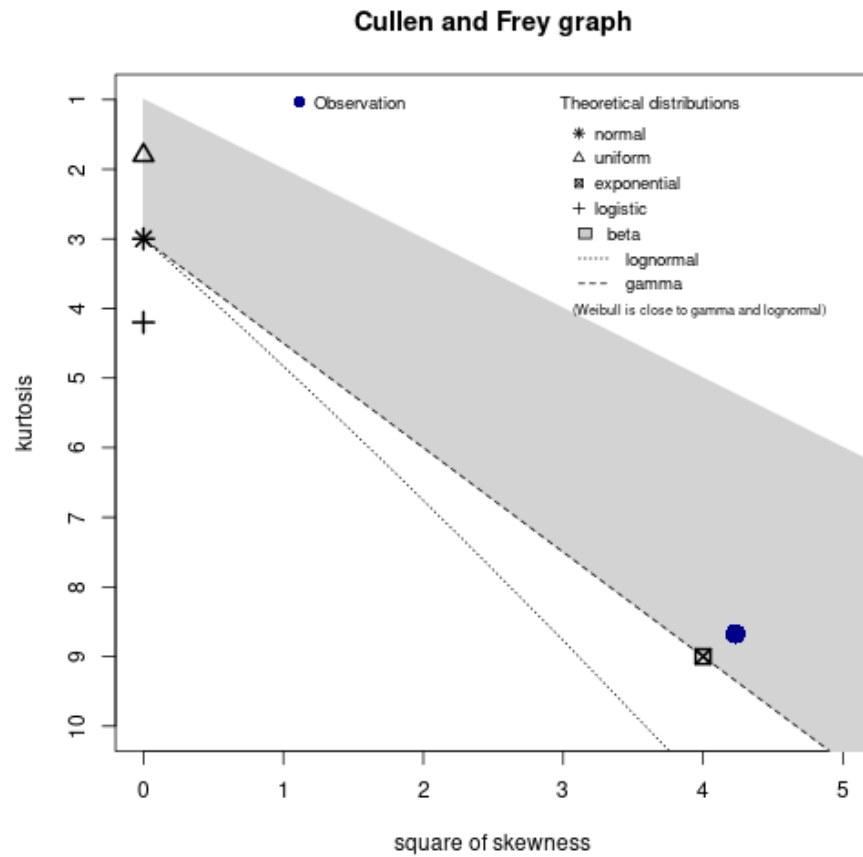
Para os dados 1:

```
descdist(dados_1, discrete = FALSE)
```



De acordo com grafico, eh possivel perceber que a amostra de dados se aproxima mais de uma distribuicao gama, ja que o ponto azul se aproxima mais da reta tracejada. Agora, para os dados 2:

```
descdist(dados_2, discrete = FALSE)
```



De acordo com grafico, eh possivel perceber que a amostra de dados se aproxima mais de uma distribuicao exponencial, ja que o ponto azul se aproxima mais do ponto marcado como exponencial.

e)

Para os dados 1:

```
fit.weibull_1$aic
## [1] 469.4881
fit.gamma_1$aic
## [1] 461.497
fit.lognormal_1$aic
## [1] 469.2617
fit.normal_1$aic
```

```
## [1] 515.0587
```

```
fit.expo_1$aic
```

```
## [1] 561.2366
```

Como o AIC do fit da gama eh o menor, percebe-se que, de fato, o ajuste para a distribuicao gama eh o melhor dentre os demais para o primeiro conjunto de dados.

Agora, para os dados 2:

```
fit.weibull_2$aic
```

```
## [1] -72.59204
```

```
fit.gamma_2$aic
```

```
## [1] -72.64928
```

```
fit.lognormal_2$aic
```

```
## [1] -33.86214
```

```
fit.normal_2$aic
```

```
## [1] 99.71861
```

```
fit.expo_2$aic
```

```
## [1] -74.49791
```

Como o AIC do fit da exponencial eh o menor, percebe-se que, de fato, o ajuste para a distribuicao exponencial eh o melhor dentre os demais para o segundo conjunto de dados.

f)

```
ks.test(dados_1, "pgamma", 3.027141, 2.032940, exact=FALSE)
```

```
##
```

```
## Asymptotic one-sample Kolmogorov-Smirnov test
```

```
##
```

```
## data: dados_1
```

```
## D = 0.045778, p-value = 0.796
```

```
## alternative hypothesis: two-sided
```

```
ks.test(dados_2, "pexp", 3.291175, exact=FALSE)
```

```
##
```

```
## Asymptotic one-sample Kolmogorov-Smirnov test
```

```
##
```

```
## data: dados_2
```

```
## D = 0.033408, p-value = 0.9789
```

alternative hypothesis: two-sided

Como os p-valores são 0.796 e 0.9789 para, respectivamente, os testes da gamma e da exponencial, podemos concluir que os ajustes são adequados, uma vez que os p-valores são superiores a 5%.

Questão 3

a)

Pelo método dos momentos temos:

$$\mu_1 = \int_0^\infty \frac{y^2 * e^{\frac{-y}{\beta}}}{\beta^2} dy = E[Y] \quad \therefore \quad \mu_1 = E[Y] = 2\beta$$
$$\mu_1 = E[Y] = 2\beta \quad \therefore \quad \beta = \mu_1/2$$

Agora,

$$m_1 = \mu = \mu_1 \quad \therefore \quad \beta = \frac{\mu}{2}$$

Além disso,

$$\mu_2 = E[Y^2] = \sigma^2 + \mu^2 = \int_0^\infty \frac{y^3 * e^{\frac{-y}{\beta}}}{\beta^2} dy = 6\beta^2$$
$$\sigma^2 + 4\beta^2 = 6\beta^2 \quad \therefore \quad \sigma^2 = 2\beta^2$$

Portanto:

$$E[\beta] = E\left[\frac{<Y>}{2}\right] = \frac{\mu}{2}$$
$$Var[\beta] = E\left[\frac{<Y>^2}{4}\right] - E\left[\frac{<Y>}{2}\right]^2 = \frac{\sigma^2/n + \mu^2}{4} - \mu^2/4 = \frac{\mu^2}{8n}$$

b)

Pelo método da máxima verossimilhança, temos:

$$L(\beta) = f(y_1) * f(y_2) \cdots * f(y_n) = y_1 * y_2 \cdots * y_n * \frac{e^{\frac{-n* <Y>}{\beta}}}{\beta^{2n}}$$
$$l(\beta) = \ln(L(\beta)) = \ln(y_1) + \ln(y_2) + \cdots + \ln(y_n) - \frac{n* <Y>}{\beta} - 2n\ln(\beta)$$
$$\frac{dl(\beta)}{d\beta} = 0 \quad \therefore \quad \frac{n* <Y>}{\beta} = 2n \quad \therefore \quad \beta_{MV} = \frac{<Y>}{2}$$

Mas,

$$E[Y] = \int_0^\infty \frac{y^2 * e^{\frac{-y}{\beta}}}{\beta^2} dy = 2\beta = <Y>$$

$$E[Y^2] = Var[Y] + \langle Y \rangle^2 = \int_0^\infty \frac{y^3 * e^{-\frac{y}{\beta}}}{\beta^2} dy = 6\beta^2 \quad \therefore \quad Var[Y] = 2\beta^2 = \frac{\langle Y \rangle^2}{2}$$

Portanto, temos:

$$E[\beta] = \langle Y \rangle \quad e \quad Var[\beta] = \frac{\langle Y \rangle^2}{2}$$

Questão 4

a)

$$\mu_{X_1} = \frac{58 + 72 + 72 + 86 + 86 + 100 + 100 + 114 + 114 + 128}{10} = 93$$

b)

$$\mu_{Y_1} = \frac{80 + 80 + 90 + 90 + 100 + 100 + 110 + 110 + 120 + 120}{10} = 100$$

c)

$$Cov[X_1, Y_1] = E[X_1 Y_1] - \mu_{X_1} * \mu_{Y_1} = 1/10 * \left(\sum_{n=1}^{10} x_n * y_n \right) - 9300$$

$$Cov[X_1, Y_1] = 9580 - 9300 = 280$$

d)

$$E[X_1^2] = \sum_{n=1}^6 x_n^2 * P(x_n) = 0.1 * 58^2 + 0.2 * 72^2 + 0.2 * 86^2 + 0.2 * 100^2 + 0.2 * 114^2 + 0.1 * 128^2$$

$$E[X_1^2] = 9090$$

e)

$$E[Y_1^2] = \sum_{n=1}^5 x_n^2 * P(x_n) = 0.2 * 80^2 + 0.2 * 90^2 + 0.2 * 100^2 + 0.2 * 110^2 + 0.2 * 120^2$$

$$E[Y_1^2] = 10200$$

f)

$$Var[X_1] = Cov[X_1^2] = E[X_1^2] - \mu_{X_1}^2$$

$$Var[X_1] = 9090 - 93 * 93 = 441$$

g)

$$Var[Y_1] = Cov[Y_1^2] = E[Y_1^2] - \mu_{Y_1}^2$$

$$Var[Y_1] = 10200 - 100 * 100 = 200$$

h)

$$\rho_{X_1, Y_1} = \frac{Cov(X_1, Y_1)}{\sigma_{X_1} * \sigma_{Y_1}}$$

$$\rho = 0.942809042$$

i) Obteve-se que a correlacao entre X2 e Y2 eh -0.942809. Com isso, observa-se que a correlacao eh igual em modulo, mas diferentes em sinal. Tal fato indica que enquanto o valor de X1 sobe eh esperado que Y1 cresca e Y2 decresca.

Questão 5

a)

$$E[X] = \sum_{n=1}^3 x_n * P(x_n) = -3/8 + 0 + 3/8 = 0$$

b)

$$E[Y] = \sum_{n=1}^3 y_n * P(y_n) = -3/8 + 0 + 3/8 = 0$$

c)

$$Cov[X, Y] = E[X, Y] - E[X]E[Y] = \sum_{n=1}^3 x_n * y_n * P_{X,Y}(x_n, y_n)$$

$$Cov[X, Y] = 1/8 + 1/8 = 1/4$$

d) Não são independentes, uma vez que a covariância não é nula, logo o coeficiente de correlação não é zero.