

2001 Tuberculose e Hepatite B

Catarina Pereira nº 12943

Raquel Machado nº16206 - raquel_mrm@hotmail.com

Vera Gomes nº16019 - vera.gomes@gmail.com

LECN

1.

Tuberculose

É uma doença infecciosa causada por um micróbio chamado "bacilo de Koch", contagiosa. Atinge principalmente os pulmões, mas pode ocorrer em outras partes do nosso corpo, como nos gânglios, rins, ossos, intestinos e meninges. A tuberculose transmite-se apenas quando o doente tosse, fala ou espirra ele espalha no ar gotas pequenas, mas muito pequenas mesmo com o micróbio da tuberculose. É assim que acontece o contágio: o micróbio da tuberculose (bacilo de Koch) penetra no organismo das pessoas pela respiração. Portanto, a tuberculose não se transmite pelo sexo, pelo sangue contaminado, pelo beijo, pelo copo, pelos talheres, pela roupa, pelo colchão... A tuberculose só se transmite pelo AR.

Fontes:

<http://www.saude.rj.gov.br/Tuberculose/perguntas.shtml>

<http://www.who.int/mediacentre/factsheets/fs104/en/>

Hepatite B

É uma inflamação do fígado, causada pelo vírus HBV (Hepatite B V), podendo haver vários tipos de hepatite A, B, C, D e E. Todos estes vírus podem causar uma doença aguda com os sintomas que duram diversas semanas incluindo a pele ficar amarelada; urina escura; fezes claras; fadiga extrema; náusea; vômitos e dor abdominal. O vírus da Hepatite B pode causar infecção crônica levando mais tarde ao desenvolvimento de cirrose e cancro do fígado. HBV é o tipo o mais sério de Hepatite viral e o único tipo que causa Hepatite crônica. Existe uma vacina disponível para a Hepatite B. Esta transmite-se por transfusões de sangue, partilha de seringas, agulhas ect, assim como relações sexuais sem protecção.

Fontes:

<http://www.abcdasaude.com.br/artigo.php?229>

<http://www.who.int/mediacentre/factsheets/fs204/en/>

2.

Para a Tuberculose

```
>dados<-read.table("2001TuberculoseHepatiteB.txt", head=T)
#Lê a tabela do ficheiro "2001TuberculoseHepatiteB.txt" que está localizado na pasta definida
no R
```

```

#A primeira linha da tabela correspondem aos títulos
> summary(dados$Tuberculose)
  Min. 1st Qu.  Median   Mean 3rd Qu.   Max.
  0.00  6.00  10.00  17.13  18.00  418.00
> L<-max(dados$Tuberculose)-min(dados$Tuberculose) #Amplitude dos dados
> L
[1] 418
> k<-ceiling(1+log(441)/log(2)) #Calcula o menor inteiro não inferior a 1+log(20)/log(2)
> k
[1] 10
> l<-L/k #Amplitude de cada classe
> l
[1] 41.8
> j=k+1
> j
[1] 11
> i<-1:j
> i
[1] 1 2 3 4 5 6 7 8 9 10 11
> quebra<-min(dados$Tuberculose)+(i-1)*l
> for(i in 1:j) {
+ quebra[i]<-min(dados$Tuberculose)+(i-1)*l } #Cálculo dos extremos de cada classe
> quebra
[1] 0.0 41.8 83.6 125.4 167.2 209.0 250.8 292.6 334.4 376.2 418.0
> Fi<-rep(0,k)
> for(i in 1:k)
+ {Fi[i]<-sum(dados$Tuberculose<=quebra[i+1])} #Frequências absolutas acumuladas
> Fi
[1] 415 434 438 438 439 440 441 441 441 442
> fi<-rep(0,k)
> fi[1]<-Fi[1]
> for(i in 2:k)
+ {fi[i]<-Fi[i]-Fi[i-1]} #Frequências absolutas
> fi
[1] 415 19 4 0 1 1 1 0 0 1
> fi_ast<-fi/Fi[k] #Frequências relativas
> fi_ast
[1] 0.938914027 0.042986425 0.009049774 0.000000000 0.002262443 0.002262443
[7] 0.002262443 0.000000000 0.000000000 0.002262443
> Fi_ast<-Fi/Fi[k] #Frequências relativas acumuladas
> Fi_ast
[1] 0.9389140 0.9819005 0.9909502 0.9909502 0.9932127 0.9954751 0.9977376
[8] 0.9977376 0.9977376 1.0000000
> cl_inf<-quebra[-(k+1)] #Extremos inferiores das classes
> cl_inf
[1] 0.0 41.8 83.6 125.4 167.2 209.0 250.8 292.6 334.4 376.2
> cl_sup<-quebra[-1] #Extremos superiores das classes

```

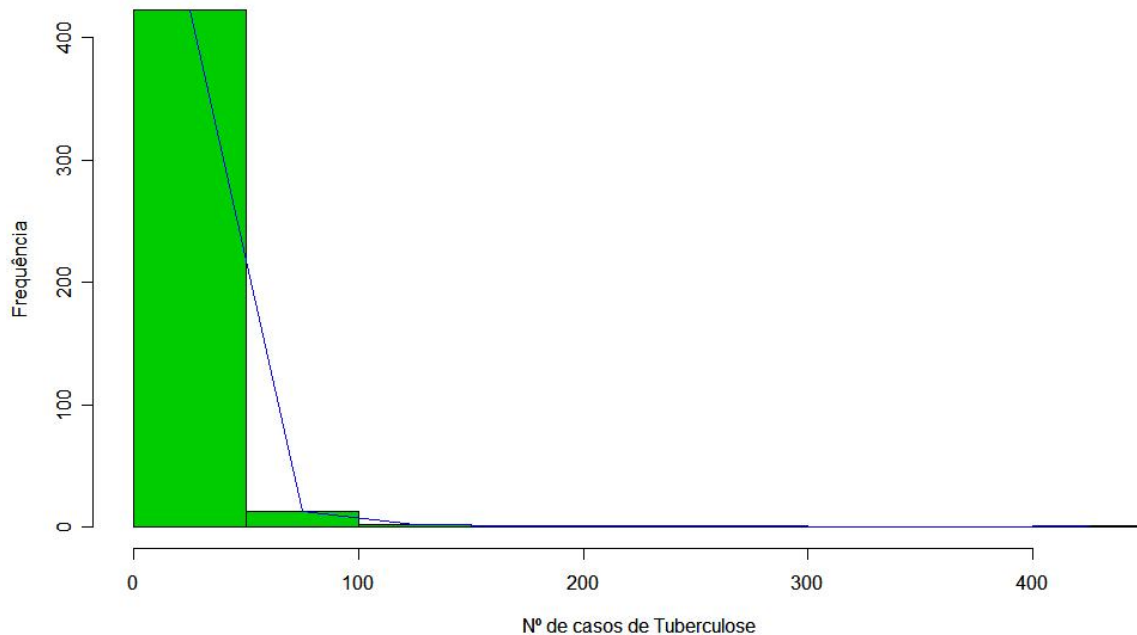
```
> cl_sup
[1] 41.8 83.6 125.4 167.2 209.0 250.8 292.6 334.4 376.2 418.0
```

Classes i	Freq. Abs. f_i	Freq. Abs. Acum. F_i	Freq. Rel. f_i^*	Freq. Rel. Acum. F_i^*
[0;41.8[415	415	0.938914027	0.9389140
[41.8;83.6[19	434	0.042986425	0.9819005
[83.6;125.4[4	438	0.009049774	0.9909502
[125.4;167.2[0	438	0.000000000	0.9909502
[167.2;209.0[1	439	0.002262443	0.9932127
[209.0;250.8[1	440	0.002262443	0.9954751
[250.8;292.6[1	441	0.002262443	0.9977376
[292.6;334.4[0	441	0.000000000	0.9977376
[334.4;376.2[0	441	0.000000000	0.9977376
[376.2;418.0[1	442	0.002262443	1.0000000

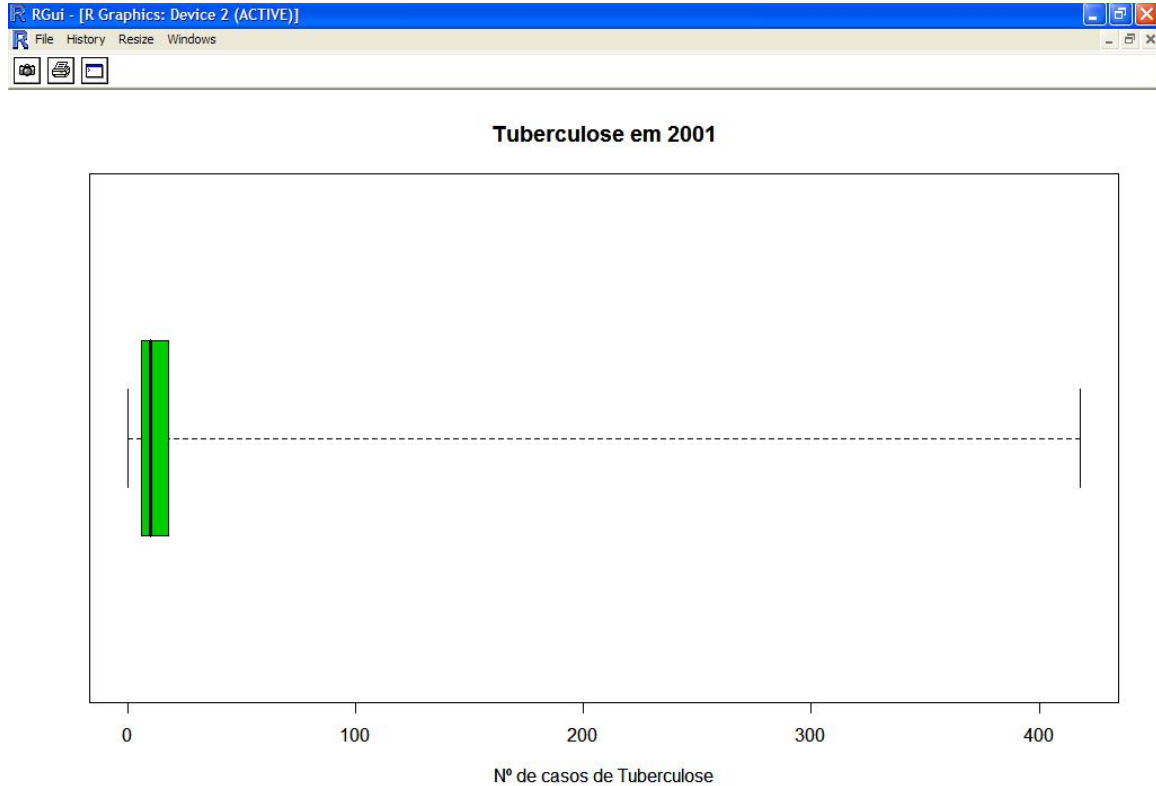
```
> lines(hist(dados$Tuberculose)$mids,hist(dados$Tuberculose, col="3", ylab="Frequência", xlab="Nº de casos de Tuberculose", main="Histograma dos dados de Tuberculose em 2001")$counts,col="blue")
```



Histograma dos dados de Tuberculose em 2001



```
>boxplot(dados$Tuberculose, range=0, horizontal=TRUE, col="3", xlab="Nº de
casos de Tuberculose", main="Tuberculose em 2001")
```



Conclusão:

Através da análise dos dados podemos concluir que em 441 concelhos da Alemanha verificou-se uma incidência de tuberculose, número de casos 418, em cerca de 50 concelhos deste país. Deste modo, conclui-se que a incidência de tuberculose não se expressa significativamente, por se verificar apenas em 11% dos concelhos. Os restantes concelhos apresentam um número de casos muito inferior à média considerada, 17.13. O diagrama caixa-e-bigode e o histograma confirmam os valores obtidos verificando-se uma acentuada assimetria esquerda dos dados.

[Para a Hepatite B](#)

```
> summary(dados$HepatiteB)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
 0.000  1.000   3.000  5.342  5.000 243.000
> L<-max(dados$HepatiteB)-min(dados$HepatiteB) #Amplitude dos dados
```

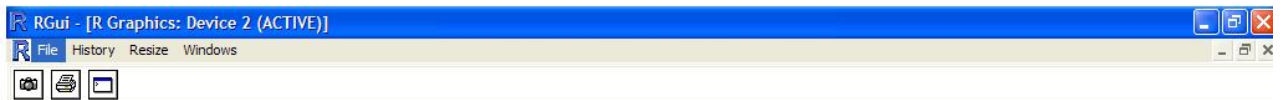
```

> L
[1] 243
> k<-ceiling(1+log(441)/log(2)) #Calcula o menor inteiro não inferior a 1+log(20)/log(2)
> k
[1] 10
> l<-L/k #Amplitude de cada classe
> l
[1] 24.3
> j=k+1
> j
[1] 11
> i<-1:j
> i
[1] 1 2 3 4 5 6 7 8 9 10 11
> quebra<-min(dados$HepatiteB)+(i-1)*l
for(i in 1:j)
{quebra[i]<-min(dados$HepatiteB)+(i-1)*l} #Cálculo dos extremos de cada classe
> quebra
[1] 0.0 24.3 48.6 72.9 97.2 121.5 145.8 170.1 194.4 218.7 243.0
> Fi<-rep(0,k)
> for(i in 1:k)
+ {Fi[i]<-sum(dados$HepatiteB<=quebra[i+1])} #Frequências absolutas acumuladas
> Fi
[1] 431 438 439 440 441 441 441 441 441 442
> fi<-rep(0,k)
> fi[1]<-Fi[1]
> for(i in 2:k)
+ {fi[i]<-Fi[i]-Fi[i-1]} #Frequências absolutas
> fi
[1] 431 7 1 1 1 0 0 0 0 1
> fi_ast<-fi/Fi[k] #Frequências relativas
> fi_ast
[1] 0.975113122 0.015837104 0.002262443 0.002262443 0.002262443 0.000000000
[7] 0.000000000 0.000000000 0.000000000 0.002262443
> Fi_ast<-Fi/Fi[k] #Frequências relativas acumuladas
> Fi_ast
[1] 0.9751131 0.9909502 0.9932127 0.9954751 0.9977376 0.9977376 0.9977376
[8] 0.9977376 0.9977376 1.0000000
> cl_inf<-quebra[-(k+1)] #Extremos inferiores das classes
> cl_inf
[1] 0.0 24.3 48.6 72.9 97.2 121.5 145.8 170.1 194.4 218.7
> cl_sup<-quebra[-1] #Extremos superiores das classes
> cl_sup
[1] 24.3 48.6 72.9 97.2 121.5 145.8 170.1 194.4 218.7 243.0

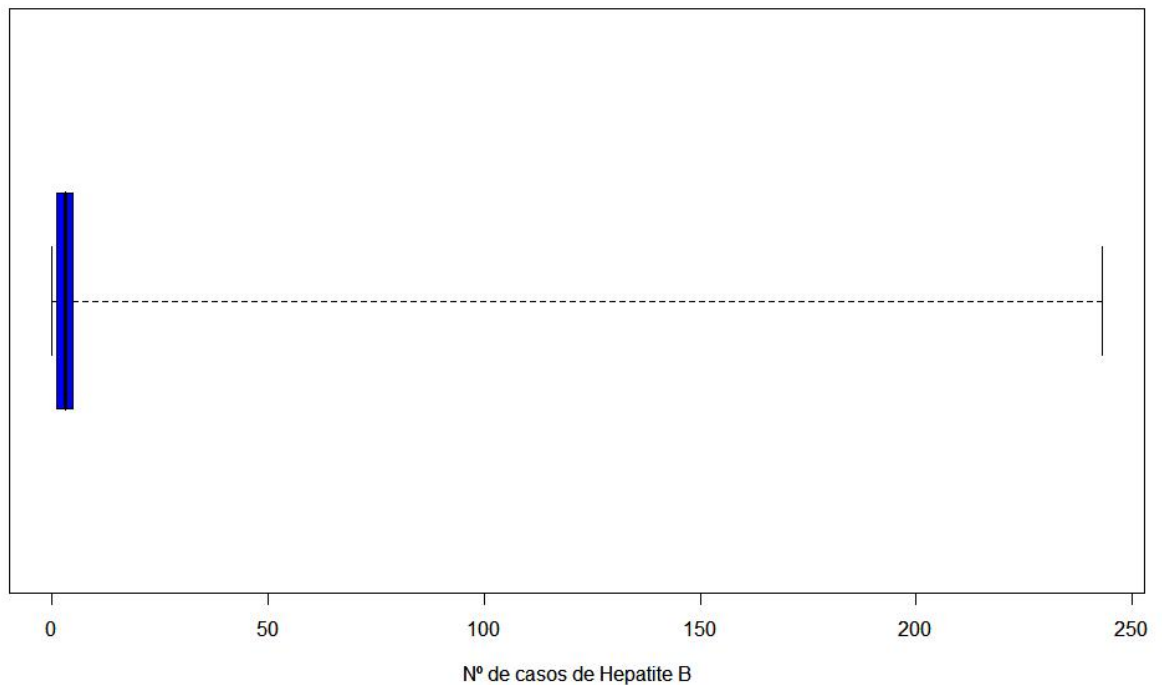
```

Classes i	Freq. Abs fi	Freq. Abs. Acum. Fi	Freq. Rel. <i>fi</i> *	Freq. Rel. Acum. <i>Fi</i> *
[0;24.3[431	431	0.975113122	0.9751131
[24.3;48.6[7	438	0.015837104	0.9909502
[48.6;72.9[1	439	0.002262443	0.9932127
[72.9;97.2[1	440	0.002262443	0.9954751
[97.2;121.5[1	441	0.002262443	0.9977376
[121.5;145.8[0	441	0.000000000	0.9977376
[145.8;170.1[0	441	0.000000000	0.9977376
[170.1;194.4[0	441	0.000000000	0.9977376
[194.4;218.7[0	441	0.000000000	0.9977376
[218.7;243.0[1	442	0.002262443	1.0000000

```
>boxplot(dados$HepatiteB, range=0, horizontal=TRUE, col="blue", xlab="Nº de casos de Hepatite B", main="Hepatite B em 2001")
```



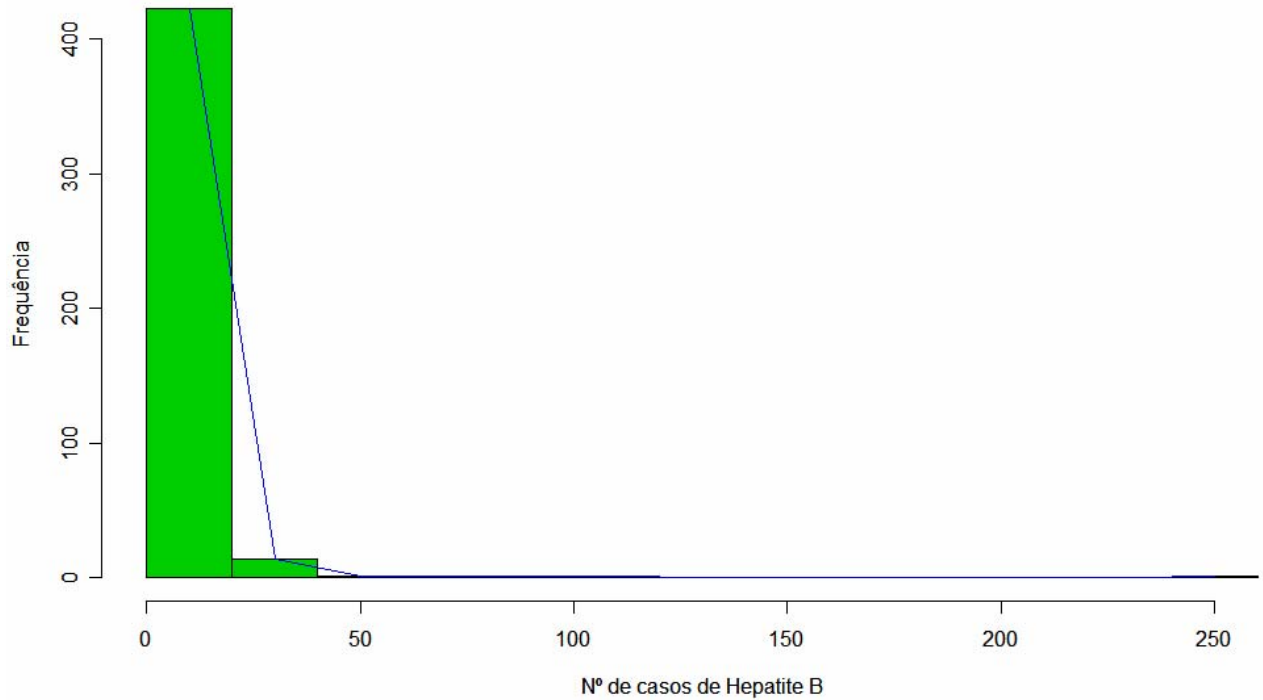
Hepatite B em 2001



```
>lines(hist(dados$HepatiteB)$mids,hist(dados$HepatiteB, col="3", ylab="Frequência", xlab="N° de casos de Hepatite B", main="Histograma dos dados de Hepatite B em 2001")$counts,col="blue")
```



Histograma dos dados de Hepatite B em 2001



Conclusão:

Através da análise dos dados podemos concluir que em 441 concelhos da Alemanha verificou-se uma incidência de Hepatite B, número de casos 243, em cerca de 25 concelhos deste país. Deste modo, conclui-se que a incidência de Hepatite B não se expressa significativamente, por se verificar apenas em 6% dos concelhos. Os restantes concelhos apresentam um número de casos muito inferior à média considerada, 5.34. O diagrama caixa-e-bigode e o histograma confirmam os valores obtidos verificando-se uma acentuada assimetria esquerda dos dados.

3.

Número médio de casos de Tuberculose e de Hepatite B em 2001 pontualmente:

```
> mean(dados$Tuberculose)
[1] 17.12896
> mean(dados$HepatiteB)
[1] 5.341629
```

Número médio de casos de Tuberculose e de Hepatite B em 2001 por intervalo de confiança:

```
> ICmedia<-function(dados,normal=TRUE,sigma=0,niv.conf=95){
+ xbar<-mean(dados)
+ n<-nrow(as.data.frame(dados)) #Dá o número de linhas, ou seja, de elementos, do ficheiro de
dados
+ if(!sigma) { #Se o sigma é desconhecido
+ sigma<-sd(dados)
+ if (normal && n < 30){ #Se a população é normal e n < 30
+ t<-qt(1-(1-niv.conf/100)/2,n-1)
+ return(list(IC=c(xbar-t*sigma/sqrt(n),xbar+t*sigma/sqrt(n)))) }
+ if (n > 30) { #Se n > 30, com população normal ou não
+ z<-qnorm(1-(1-niv.conf/100)/2,0,1)
+ return(list(IC=c(xbar-z*sigma/sqrt(n),xbar+z*sigma/sqrt(n)))) }
+ else return("Não se pode determinar IC")
+ }
+ else{ #Sigma conhecido
+ if (normal) { #População normal
+ z<-qnorm(1-(1-niv.conf/100)/2,0,1)
+ return(list(IC=c(xbar-z*sigma/sqrt(n),xbar+z*sigma/sqrt(n))))
+ }
+ else { #População não normal
+ if (n > 30) { #n > 30
+ z<-qnorm(1-(1-niv.conf/100)/2,0,1)
+ return(list(IC=c(xbar-z*sigma/sqrt(n),xbar+z*sigma/sqrt(n)))) }
+ else return("Não se pode determinar IC") }
+ }
+ }
> ICmedia(dados)
$IC
Tuberculose HepatiteB Tuberculose HepatiteB
14.345800 3.985070 19.912119 6.698188
```

Para o cálculo do intervalo de confiança do número médio de casos, tanto de Tuberculose como de Hepatite B, foi considerada uma população com distribuição normal e um nível de confiança de 95%.

- Para a tuberculose:

$x=17,12896$
 $\mu=95\%..(14.345800;19.912119)IC$

Existe uma probabilidade de 95% do valor médio de casos de Tuberculose em 2001 μ estar no intervalo [14.3;19.9].

- Para a Hepatite B:

$x=5.341629$
 $95\%..(3.985070;6.698188)IC\mu$

Existe uma probabilidade de 95% do valor médio de casos de Hepatite B em 2001 μ estar no intervalo [3.9;6.7].

4.

$$H_0: \mu_T = \mu_H \text{ vs } H_1: \mu_T \neq \mu_H$$

$$H_0: \mu_T - \mu_H = 0 \text{ vs } H_1: \mu_T - \mu_H \neq 0$$

```
> THmedias<-function(dados,normal=TRUE,sigma=0,niv.conf=95){
+ xm1<-mean(dados$Tuberculose)
+ xm2<-mean(dados$HepatiteB)
+ n1<-nrow(as.data.frame(dados$Tuberculose))
+ n2<-nrow(as.data.frame(dados$HepatiteB))
+ sigma1<-sd(dados$Tuberculose)
+ sigma2<-sd(dados$HepatiteB)
+ c<-qnorm(1-(1-niv.conf/100)/2,0,1)
+ z<-((xm1-xm2)/sqrt(((sigma1^2)/n1)+((sigma2^2)/n2)))
+ if (z<(-c) || z>c) { #Se o valor observado (z) pertencer a (-oo,-c)U(c,+oo)
+ return("A hipótese das médias serem iguais (H0) é rejeitada, pois a um nível de significância de 5%
o valor observado (z) pertence à região de rejeição.") }
+ else { #Se o valor observado (z) não pertencer a (-oo,-c)U(c,+oo)
+ return("A hipótese das médias serem iguais (H0) não é rejeitada, pois a um nível de significância de
5% o valor observado (z) não pertence à região de rejeição.")
+ }
+ }
> THmedias(dados)
[1] "A hipótese das médias serem iguais (H0) é rejeitada, pois a um nível de significância de 5% o
valor observado (z) pertence à região de rejeição."
```

Para confirmar o resultado:

```
> #Definem-se as variáveis no R.  
> niv.conf=95  
> xm1<-mean(dados$Tuberculose)  
> xm2<-mean(dados$HepatiteB)  
> n1<-nrow(as.data.frame(dados$Tuberculose))  
> n2<-nrow(as.data.frame(dados$HepatiteB))  
> sigma1<-sd(dados$Tuberculose)  
> sigma2<-sd(dados$HepatiteB)  
> c<-qnorm(1-(1-niv.conf/100)/2,0,1)  
> c  
[1] 1.959964  
> z<-((xm1-xm2)/sqrt(((sigma1^2)/n1)+((sigma2^2)/n2)))  
> z  
[1] 7.461733
```

Como se pode verificar o resultado confirma-se pois o valor observado, $z=7.461733$, encontra-se dentro da região de rejeição: $R \equiv (-\infty; -c) \cup (c; +\infty) \equiv (-\infty; -1,96) \cup (1,96; +\infty)$.

Concluo então que a hipótese das médias serem iguais (H_0), numa população com distribuição normal, é rejeitada pois a um nível de significância de 5%, o valor observado (z) pertence à região de rejeição, o que se confirma pelo calculo das médias amostrais.

Recta de regressão linear:

```
> y<-(dados$Tuberculose)  
> x<-(dados$HepatiteB)  
> lm(y~x)$coefficients #Dá os parâmetros da recta de regressão linear.  
(Intercept)      x  
 9.668404  1.396682
```

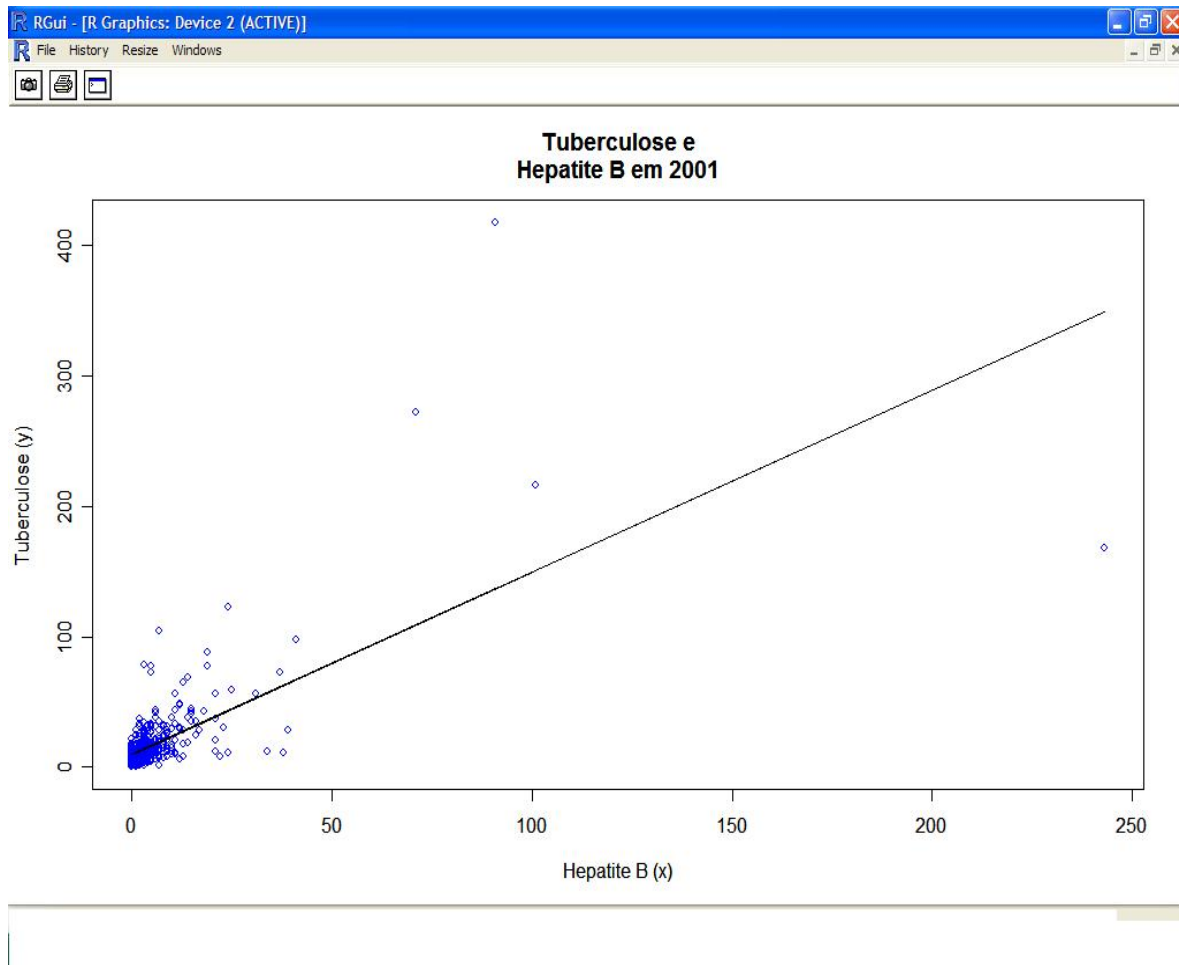
A recta de regressão linear é então dada por:

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$$
$$\bar{Y} = 9.668404 + 1.396682x$$

4.c)

Gráfico dos pontos amostrais da amostra

```
> plot(x,y, col="blue", xlab="Hepatite B (x)", ylab="Tuberculose (y)", main="Tuberculose e  
+ Hepatite B em 2001")
```



```
> lines(x,lm(y~x)$fitted)
> n<-nrow(as.data.frame(dados)) #Dá o número de linhas, ou seja, de elementos, do ficheiro de
dados
> n
[1] 442
> R2<-(1-sum((lm(y~x)$residuals)^2))/(sum(y^2)-n*mean(y)^2)
> R2
[1] 0.4634426
```

4. d)

A qualidade do ajuste é dada pelo coeficiente de determinação. Quanto mais perto de 1 estiver este coeficiente melhor é o ajuste, pelo contrário quanto mais perto estiver de 0 pior é o ajuste. Como o valor da qualidade é 0.5, pode-se afirmar que o ajuste não é muito bom.

4.e)

Teste da hipótese do declive da recta de regressão (β_1) ser nulo:

$$H_0: \hat{\beta}_1 = 0 \text{ vs } H_1: \hat{\beta}_1 \neq 0$$

```
> THB1<-function(dados, niv.conf=95){
+ n<-nrow(as.data.frame(dados)) #Dá o número de linhas, ou seja, de elementos, do ficheiro de
dados
+ y<-(dados$Tuberculose)
+ x<-(dados$HepatiteB) #Valores de Hepatite B
+ B1<-((sum(dados$Tuberculose*dados$HepatiteB)-n*(mean(x))*(mean(y))) /
+ (sum(x^2)-n*(mean(x)^2)))
+ sigmaH<-sd(x)
+ c<-qt(1-(1-niv.conf/100)/2,n-2)
+ t<-B1/sqrt((sigmaH^2)/(sum(x^2)-n*mean(x)^2))
+ if (t<(-c) || t>c) { #Se o valor observado (t) pertencer a (-oo,-c)U(c,+oo).
+ return("A hipótese de o declive (B1) da recta de regressão ser nulo (H0) é rejeitada, pois a um nível
de significância de 5% o valor observado (t) pertence à região de rejeição.") }
+ else { #Se o valor observado (t) não pertencer a (-oo,-c)U(c,+oo).
+ return("A hipótese de o declive (B1) da recta de regressão ser nulo (H0) não é rejeitada, pois a um
nível de significância de 5% o valor observado (t) não pertence à região de rejeição.")
+ }
+ }
> THB1(dados)
[1] "A hipótese de o declive (B1) da recta de regressão ser nulo (H0) é rejeitada, pois a um nível de
significância de 5% o valor observado (t) pertence à região de rejeição."
```

Para confirmar o resultado:

```
> #Definem-se as variáveis no R.
> niv.conf=95
> n<-nrow(as.data.frame(dados)) #Dá o número de linhas, ou seja, de elementos, do ficheiro de
dados
> n
[1] 442
> y<-(dados$Tuberculose)
> x<-(dados$HepatiteB) #Valores de Hepatite B
```

```

> B1<-((sum(dados$Tuberculose*dados$HepatiteB)- n*(mean(x))*(mean(y))) /
+ (sum(x^2)-n*(mean(x)^2)))
> sigmaH<-sd(x)
> c<-qt(1-(1-niv.conf/100)/2,n-2)
> c
[1] 1.96537
> t<-B1/sqrt((sigmaH^2)/(sum(x^2)-n*mean(x)^2))
> t
[1] 29.33032

```

Como se pode verificar, o resultado confirma-se pois o valor observado, $t=29.33032$, encontra-se dentro da região de rejeição: $R \in (-\infty; -c) \cup (c; +\infty) \in (-\infty; -1,96) \cup (1,96; +\infty)$.

Concluimos então que a hipótese de o declive (B1) da recta de regressão ser nulo (H0), utilizando uma distribuição t-student, é rejeitada, pois a um nível de significância de 5% o valor observado (t) pertence à região de rejeição.

Comparando com o valor calculado na recta de regressão linear este resultado encontra-se correcto, já que: $\bar{Y} = 6.33 + 3.34x$ e $\hat{1} = 3.34\beta^1$

Para que serve a recta que ajustou?

A recta ajustada serve para estabelecer uma relação entre a variável y e x, neste caso entre os casos de tuberculose e os casos de hepatite B em 441 “concelhos” da Alemanha no ano de 2001.

O conhecimento das relações entre estas variáveis permite prever o valor de variáveis em função de outras.

Neste caso, apesar da qualidade do ajuste não ser muito boa, quanto maior é o número de casos de Hepatite B, maior é o número de casos de Tuberculose.