

The results below are generated from an R script.

```
# Pacotes ----
if (!require("pacman")) install.packages("pacman")

## Loading required package: pacman

pacman::p_load(
  readxl,
  tidyverse,
  cowplot,
  mdscore,
  AICcmodavg,
  questionr,
  mlpack,
  ResourceSelection,
  lmtest,
  car,
  stats,
  knitr,
  pROC,
  ROCit,
  labelled,
  compareGroups,
  arm,
  performance
)

# Dados de análise e treino dos modelos ----
df <- read_excel("arquivos/Amostra_g06_Bruno_Rafael.xlsx")
colnames(df) <- c("ID", # Identificação do paciente
                  "resultado_radiografia", # 0 = negativo / 1 - positivo
                  "estagio_tumor", # 0 - menos grave / 1 - mais grave
                  "nivel_fosfatase_acida", # x100
                  "envolvimento_nodal" # 0 - não / 1 - sim
                  )

var_label(df) <- list(
  resultado_radiografia = "Resultado da Radiografia",
  estagio_tumor         = "Estágio do Tumor",
  nivel_fosfatase_acida = "Nível da Fosfatase Ácida",
  envolvimento_nodal    = "Envolvimento Nodal (x100)"
)

# Dados de validação do modelo (teste) ----
teste <- read_excel("arquivos/Amostra_VALIDACAO.xlsx")
colnames(teste) <- c("ID", # Identificação do paciente
                    "resultado_radiografia", # 0 = negativo / 1 - positivo
                    "estagio_tumor", # 0 - menos grave / 1 - mais grave
                    "nivel_fosfatase_acida", # x100
                    "envolvimento_nodal" # 0 - não / 1 - sim
                    )
```

```
# Parte 1) - Análise exploratória ----

dados = df %>%
  mutate(resultado_radiografia = case_when(
    resultado_radiografia == "0" ~ "Negativo",
    resultado_radiografia == "1" ~ "Positivo"
  ), estagio_tumor = case_when(
    estagio_tumor == "0" ~ "- grave",
    estagio_tumor == "1" ~ "+ grave"
  ), envolvimento_nodal = case_when(
    envolvimento_nodal == "0" ~ "Não",
    envolvimento_nodal == "1" ~ "Sim"
  ), nivel_fosfatase_acida = nivel_fosfatase_acida)

dados = dados[,2:5]
dados$envolvimento_nodal = factor(dados$envolvimento_nodal)
dados$estagio_tumor = factor(dados$estagio_tumor)
dados$resultado_radiografia = factor(dados$resultado_radiografia)

compareGroups(envolvimento_nodal ~ . ,
  data=dados,
  method = c(nivel_fosfatase_acida=NA), alpha= 0.05 # Testando a normalidade para decidir o
)

##
##
## ----- Summary of results by groups of 'envolvimento_nodal'-----
##
##
##      var                N  p.value method                selection
## 1 resultado_radiografia    102 <0.001** categorical          ALL
## 2 estagio_tumor            102 <0.001** categorical          ALL
## 3 Nível da Fosfatase Ácida 102 <0.001** continuous non-normal ALL
## -----
## Signif. codes:  0 '**' 0.05 '*' 0.1 '.' 1

tabela = compareGroups(envolvimento_nodal ~ . ,
  data=dados,
  method = c(nivel_fosfatase_acida=NA), alpha= 0.05 # Testando a normalidade para decidir o
)

createTable(tabela, show.ratio=TRUE)

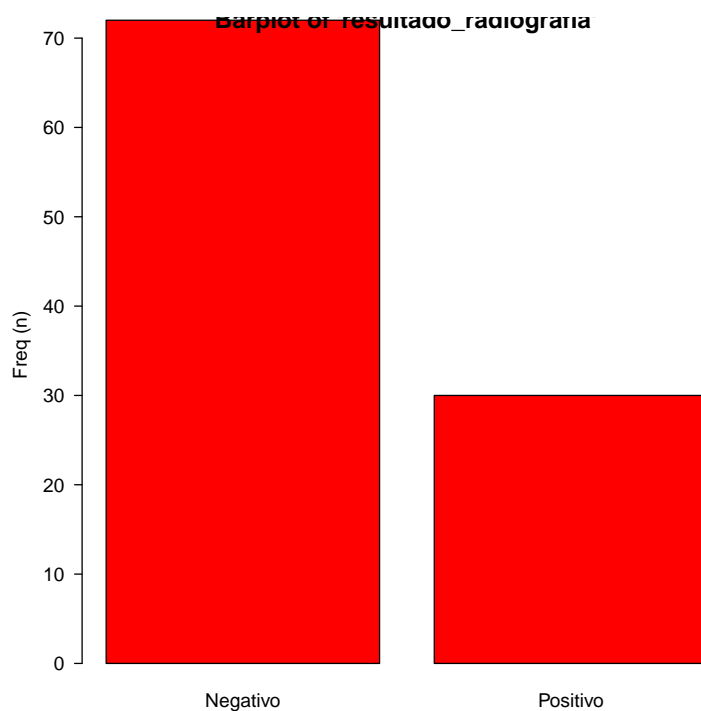
##
## -----Summary descriptives table by 'envolvimento_nodal'-----
##
## -----
##                Não                Sim                OR                p.ratio p.overall
##                N=60                N=42
## -----
## resultado_radiografia:                                     <0.001
##   Negativo          52 (86.7%)          20 (47.6%)          Ref.          Ref.
##   Positivo           8 (13.3%)          22 (52.4%)          6.92 [2.72;19.2] <0.001
## estagio_tumor:
##   - grave           39 (65.0%)           8 (19.0%)          Ref.          Ref.
##                                     <0.001
```

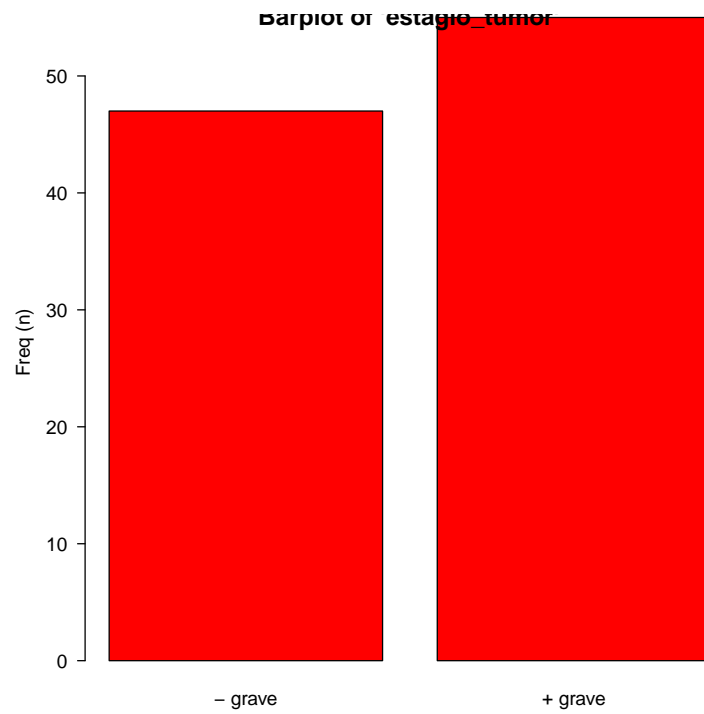
```
##      + grave                21 (35.0%)        34 (81.0%)    7.63 [3.09;20.8] <0.001
## Nível da Fosfatase Ácida 53.0 [50.0;63.2] 74.0 [67.0;83.5] 1.01 [1.00;1.03] 0.068 <0.001
## -----
```

```
descrTable(dados)
```

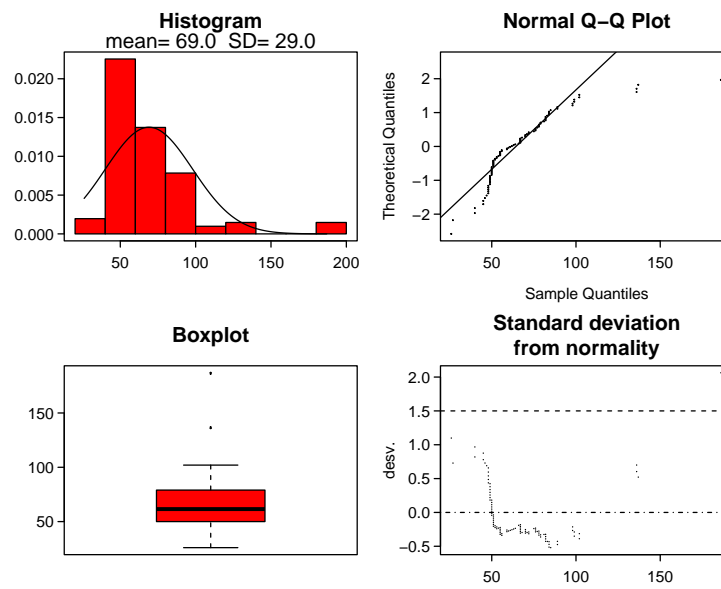
```
##
## -----Summary descriptives table -----
##
## -----
##                      [ALL]      N
##                      N=102
## -----
## resultado_radiografia:                102
##      Negativo                72 (70.6%)
##      Positivo                30 (29.4%)
## estagio_tumor:                102
##      - grave                47 (46.1%)
##      + grave                55 (53.9%)
## Nível da Fosfatase Ácida 69.0 (29.0) 102
## envolvimento_nodal:                102
##      Não                60 (58.8%)
##      Sim                42 (41.2%)
## -----
```

```
plot(tabela)
```





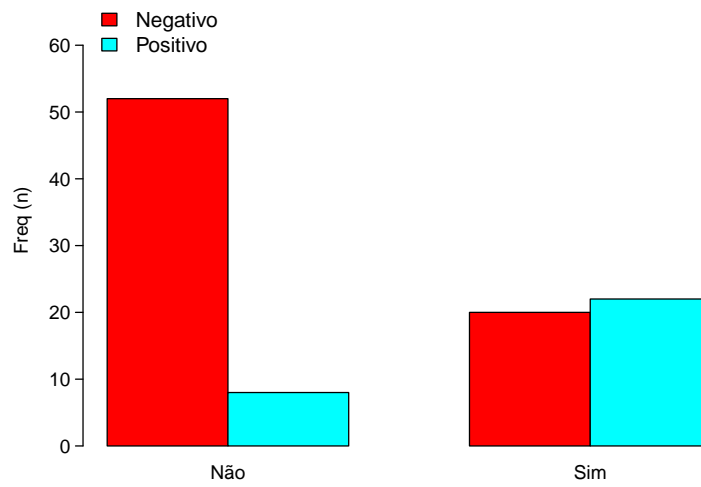
Normality plots of 'Nível da Fosfatase Ácida'



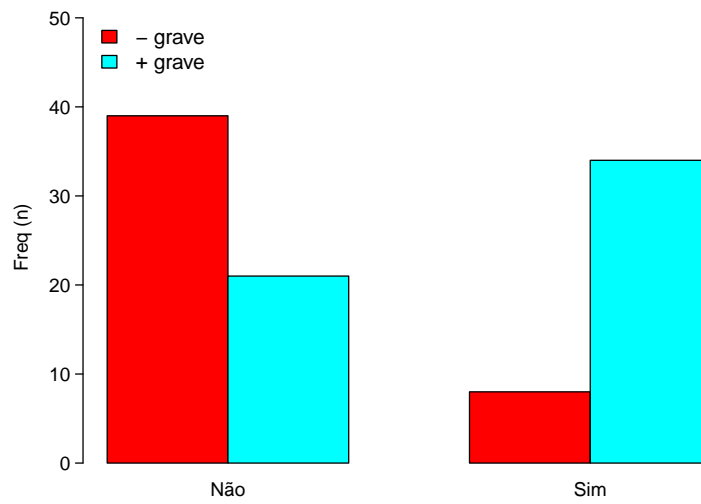
Shapiro-Wilks p-value: <0.001

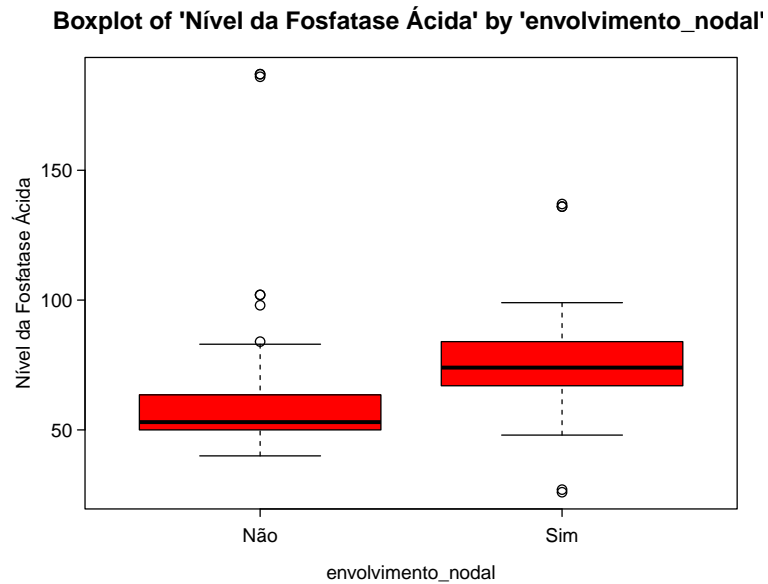
```
plot(tabela, bivar=TRUE)
```

Barplot of 'resultado_radiografia' by 'envolvimento_nodal'



Barplot of 'estagio_tumor' by 'envolvimento_nodal'



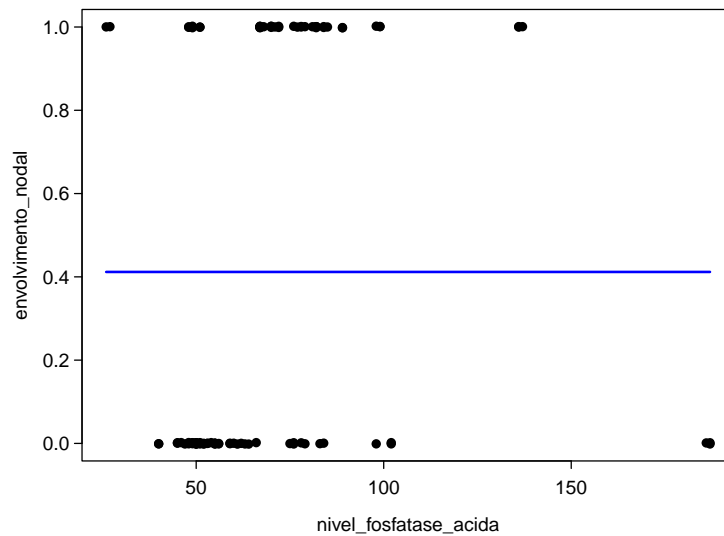


```
# rm(dados,tabela)
# Parte 2) Regressão logística: nivel_fosfatase_acida EXPLICANDO envolvimento_nodal ----

# Somente intercepto
fit0 <- glm(envolvimento_nodal ~ 1,
            family=binomial(link=logit),
            data=df)

plot(
  jitter(envolvimento_nodal, 0.01) ~ nivel_fosfatase_acida,
  xlab = "nivel_fosfatase_acida",
  ylab = "envolvimento_nodal",
  data = df,
  pch = 16
)

curve(
  predict(fit0, data.frame(nivel_fosfatase_acida = x), type = "resp"),
  add = T,
  col = "blue",
  lwd = 2
)
```



```
get_confint <- function(fit, variavel, digitos){
  if(!is.null(dim(confint(fit)))){
    round(exp(confint(fit)[variavel, ]), digitos)
  } else {
    round(confint(fit), digitos) |>
      str_c( collapse = "; ")
  }
}

fit0_pred <- predict(fit0, type="response", se.fit=TRUE)
fit0_fit <- fit0_pred$fit[1]
fit0_se <- fit0_pred$se.fit[1]

# Intervalo de confiança para a probabilidade
Lb_fit0 <- fit0_fit - qnorm(0.975) * fit0_se
Ub_fit0 <- fit0_fit + qnorm(0.975) * fit0_se

medidas0 <- as.data.frame(cbind(fit0$deviance, fit0$AIC, BIC(fit0),
                                logLik(fit0)[1]))
colnames(medidas0) <- c("Deviance", "AIC", "BIC", "Log Likelihood")
medidas0
```

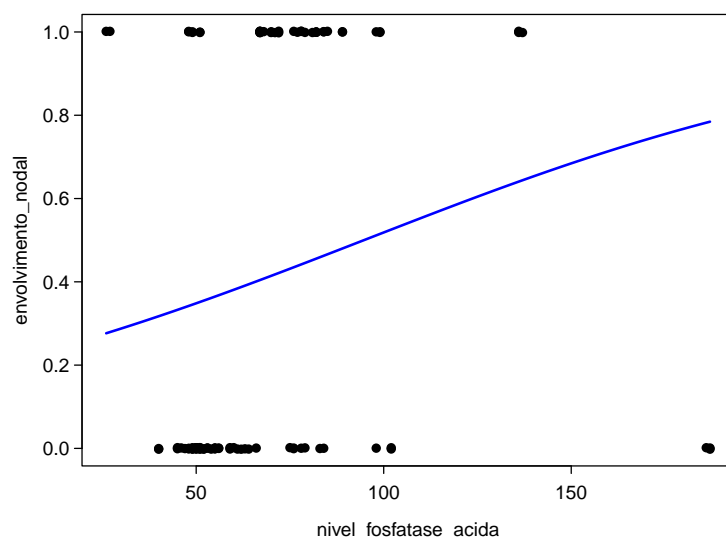
```
##      Deviance      AIC      BIC Log Likelihood
## 1 138.2089 140.2089 142.8338      -69.10443

# fit1: somente nivel_fosfatase_acida

fit1 <- glm(envolvimento_nodal ~ nivel_fosfatase_acida,
            family=binomial(link=logit),
            data=df)

plot(jitter(envolvimento_nodal,0.01) ~ nivel_fosfatase_acida, xlab="nivel_fosfatase_acida",
     ylab="envolvimento_nodal",
     data=df, pch=16)

curve(predict(fit1, data.frame(nivel_fosfatase_acida=x),type="resp"),
      add=T, col="blue", lwd=2)
```



```
summary(fit1); confint(fit1)

##
## Call:
## glm(formula = envolvimento_nodal ~ nivel_fosfatase_acida, family = binomial(link = logit),
##      data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.326089   0.564939  -2.347   0.0189 *
## nivel_fosfatase_acida  0.014003   0.007662   1.828   0.0676 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```



```
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 138.21  on 101  degrees of freedom
## Residual deviance: 134.45  on 100  degrees of freedom
## AIC: 138.45
##
## Number of Fisher Scoring iterations: 4

## Waiting for profiling to be done...

##              2.5 %      97.5 %
## (Intercept)      -2.5152009435 -0.26824134
## nivel_fosfatase_acida -0.0001477485  0.03055407

hoslem_fit1 <- ResourceSelection::hoslem.test(fit1$y, fitted(fit1), g = 10)

# DescTools::HosmerLemeshowTest(fitted(fit1), fit1$y)
# H_0) Valores observados e valores esperados são iguais para diferentes níveis de nivel_fosfatase_acida
# H_1) c.c.
# O teste de H-L rejeita a hipótese nula, portanto indica que o modelo não é adequado.

# Analisando a influência de cada valor no resultado da regressão logística
stats::influence.measures(fit1)

## Influence measures of
## glm(formula = envolvimento_nodal ~ nivel_fosfatase_acida, family = binomial(link = logit),
##
##      dfb.1_ dfb.nv__ dffit cov.r cook.d   hat inf
## 1  -0.066435  0.03918 -0.0913  1.02 0.00357 0.0120
## 2  -0.080086  0.05520 -0.0963  1.02 0.00391 0.0142
## 3  -0.075694  0.05003 -0.0945  1.02 0.00378 0.0134
## 4  -0.080086  0.05520 -0.0963  1.02 0.00391 0.0142
## 5  -0.082223  0.05772 -0.0973  1.02 0.00398 0.0147
## 6  -0.088401  0.06503 -0.1005  1.02 0.00421 0.0161
## 7  -0.051337  0.02158 -0.0887  1.02 0.00342 0.0106
## 8  -0.068810  0.04196 -0.0920  1.02 0.00361 0.0123
## 9   0.049906 -0.00754  0.1184  1.00 0.00764 0.0102
## 10 -0.086381  0.06264 -0.0994  1.02 0.00413 0.0156
## 11 -0.082223  0.05772 -0.0973  1.02 0.00398 0.0147
## 12 -0.080086  0.05520 -0.0963  1.02 0.00391 0.0142
## 13 -0.003751 -0.03330 -0.1015  1.01 0.00470 0.0116
## 14  0.013363 -0.05287 -0.1122  1.02 0.00585 0.0134
## 15  0.071230 -0.11862 -0.1606  1.02 0.01267 0.0231
## 16 -0.075694  0.05003 -0.0945  1.02 0.00378 0.0134
## 17 -0.072198  0.11808  0.1575  1.02 0.01184 0.0240
## 18  0.565103 -0.66499 -0.6818  1.13 0.36437 0.1460  *
## 19 -0.170422  0.21638  0.2329  1.08 0.02302 0.0709  *
## 20 -0.099725  0.07852 -0.1073  1.03 0.00474 0.0196
## 21 -0.080086  0.05520 -0.0963  1.02 0.00391 0.0142
## 22 -0.080086  0.05520 -0.0963  1.02 0.00391 0.0142
## 23 -0.068810  0.04196 -0.0920  1.02 0.00361 0.0123
## 24 -0.059069  0.03058 -0.0896  1.02 0.00347 0.0112
## 25  0.121575 -0.08214  0.1488  1.00 0.01306 0.0138
## 26  0.130932 -0.09192  0.1550  1.00 0.01431 0.0147
```

## 27	-0.084321	0.06020	-0.0984	1.02	0.00405	0.0151	
## 28	0.088313	-0.13792	-0.1769	1.02	0.01561	0.0267	
## 29	-0.010294	-0.02579	-0.0981	1.01	0.00437	0.0111	
## 30	-0.018868	0.06350	0.1279	1.01	0.00828	0.0139	
## 31	0.004448	0.03948	0.1203	1.01	0.00751	0.0116	
## 32	0.037175	0.00566	0.1171	1.00	0.00737	0.0102	
## 33	0.049906	-0.00754	0.1184	1.00	0.00764	0.0102	
## 34	-0.011217	0.05563	0.1250	1.01	0.00797	0.0130	
## 35	0.049906	-0.00754	0.1184	1.00	0.00764	0.0102	
## 36	0.028824	0.01430	0.1170	1.00	0.00730	0.0104	
## 37	-0.037450	0.08257	0.1366	1.01	0.00925	0.0165	
## 38	-0.082223	0.05772	-0.0973	1.02	0.00398	0.0147	
## 39	-0.068810	0.04196	-0.0920	1.02	0.00361	0.0123	
## 40	-0.073439	0.04738	-0.0936	1.02	0.00372	0.0130	
## 41	-0.084321	0.06020	-0.0984	1.02	0.00405	0.0151	
## 42	-0.090383	0.06738	-0.1016	1.02	0.00429	0.0167	
## 43	-0.048677	0.01849	-0.0886	1.02	0.00343	0.0105	
## 44	-0.071144	0.04469	-0.0928	1.02	0.00366	0.0126	
## 45	-0.053955	0.02462	-0.0889	1.02	0.00343	0.0108	
## 46	0.045635	-0.00311	0.1178	1.00	0.00753	0.0102	
## 47	-0.080086	0.05520	-0.0963	1.02	0.00391	0.0142	
## 48	-0.077910	0.05264	-0.0954	1.02	0.00384	0.0138	
## 49	-0.000414	-0.03712	-0.1034	1.01	0.00490	0.0119	
## 50	0.016915	-0.05693	-0.1147	1.02	0.00613	0.0139	
## 51	-0.077910	0.05264	-0.0954	1.02	0.00384	0.0138	
## 52	-0.010294	-0.02579	-0.0981	1.01	0.00437	0.0111	
## 53	-0.068872	0.11469	0.1553	1.02	0.01155	0.0231	
## 54	0.559232	-0.65862	-0.6755	1.13	0.35537	0.1448	*
## 55	-0.172378	0.21828	0.2344	1.08	0.02328	0.0724	*
## 56	-0.007346	0.05164	0.1237	1.01	0.00783	0.0126	
## 57	-0.077910	0.05264	-0.0954	1.02	0.00384	0.0138	
## 58	-0.066435	0.03918	-0.0913	1.02	0.00357	0.0120	
## 59	-0.056533	0.02762	-0.0893	1.02	0.00345	0.0110	
## 60	0.135640	-0.09684	0.1582	1.00	0.01499	0.0151	
## 61	-0.082223	0.05772	-0.0973	1.02	0.00398	0.0147	
## 62	-0.045977	0.01536	-0.0886	1.02	0.00344	0.0103	
## 63	-0.022647	0.06738	0.1295	1.01	0.00845	0.0143	
## 64	0.008438	0.03537	0.1194	1.01	0.00743	0.0113	
## 65	0.032985	0.00999	0.1170	1.00	0.00732	0.0103	
## 66	0.000487	0.04357	0.1213	1.01	0.00760	0.0119	
## 67	0.037175	0.00566	0.1171	1.00	0.00737	0.0102	
## 68	0.049906	-0.00754	0.1184	1.00	0.00764	0.0102	
## 69	-0.011217	0.05563	0.1250	1.01	0.00797	0.0130	
## 70	0.049906	-0.00754	0.1184	1.00	0.00764	0.0102	
## 71	0.238015	-0.20438	0.2428	1.01	0.03971	0.0290	
## 72	-0.013501	-0.02211	-0.0966	1.01	0.00422	0.0109	
## 73	-0.072198	0.11808	0.1575	1.02	0.01184	0.0240	
## 74	0.565103	-0.66499	-0.6818	1.13	0.36437	0.1460	*
## 75	-0.170422	0.21638	0.2329	1.08	0.02302	0.0709	*
## 76	-0.011217	0.05563	0.1250	1.01	0.00797	0.0130	
## 77	-0.080086	0.05520	-0.0963	1.02	0.00391	0.0142	
## 78	-0.080086	0.05520	-0.0963	1.02	0.00391	0.0142	
## 79	-0.099725	0.07852	-0.1073	1.03	0.00474	0.0196	
## 80	-0.068810	0.04196	-0.0920	1.02	0.00361	0.0123	

```
## 81 -0.059069 0.03058 -0.0896 1.02 0.00347 0.0112
## 82 0.135640 -0.09684 0.1582 1.00 0.01499 0.0151
## 83 0.121575 -0.08214 0.1488 1.00 0.01306 0.0138
## 84 0.130932 -0.09192 0.1550 1.00 0.01431 0.0147
## 85 -0.084321 0.06020 -0.0984 1.02 0.00405 0.0151
## 86 0.088313 -0.13792 -0.1769 1.02 0.01561 0.0267
## 87 -0.040450 0.00895 -0.0889 1.02 0.00348 0.0102
## 88 -0.018868 0.06350 0.1279 1.01 0.00828 0.0139
## 89 0.012458 0.03122 0.1187 1.01 0.00737 0.0111
## 90 0.037175 0.00566 0.1171 1.00 0.00737 0.0102
## 91 0.004448 0.03948 0.1203 1.01 0.00751 0.0116
## 92 -0.011217 0.05563 0.1250 1.01 0.00797 0.0130
## 93 0.049906 -0.00754 0.1184 1.00 0.00764 0.0102
## 94 0.028824 0.01430 0.1170 1.00 0.00730 0.0104
## 95 -0.037450 0.08257 0.1366 1.01 0.00925 0.0165
## 96 0.243009 -0.20965 0.2474 1.01 0.04146 0.0298
## 97 -0.082223 0.05772 -0.0973 1.02 0.00398 0.0147
## 98 -0.068810 0.04196 -0.0920 1.02 0.00361 0.0123
## 99 -0.077910 0.05264 -0.0954 1.02 0.00384 0.0138
## 100 -0.073439 0.04738 -0.0936 1.02 0.00372 0.0130
## 101 -0.077910 0.05264 -0.0954 1.02 0.00384 0.0138
## 102 -0.090383 0.06738 -0.1016 1.02 0.00429 0.0167

# teste verossimilhanca
t1 <- lr.test(fit0,fit1)
# O teste de razão de verossimilhança não rejeita (por pouco) a hipótese nula, portanto o modelo fit0 é

# teste de wald

thetahat <- fit1$coefficients
vcov1 <- vcov(fit1)
LL <- rbind(c(0,1))

WaldTest = function(L,thetahat,Vn,h=0) {
  WaldTest = numeric(3)
  names(WaldTest) = c("W","df","p-value")
  r = dim(L)[1]
  W = t(L%*%thetahat-h) %*% solve(L%*%Vn%*%t(L)) %*%
    (L%*%thetahat-h)
  W = as.numeric(W)
  pval = 1-pchisq(W,r)
  WaldTest[1] = W; WaldTest[2] = r; WaldTest[3] = pval
  WaldTest}

# teste score
score1<-anova(fit0,fit1, test="Rao")
resultados <- cbind(t1$LR,1,t1$pvalue)
resultados2 <- matrix(WaldTest(LL,thetahat,vcov1), ncol=3)
resultados3 <- cbind(anova(fit0,fit1, test="Rao")[2,4],
                     anova(fit0,fit1, test="Rao")[2,3],
                     anova(fit0,fit1, test="Rao")[2,6])

testes1 <- rbind(resultados,resultados2,resultados3)
```

```

rownames(testes1) <- c("Razao de Verossimilhanca", "Wald", "Score")
colnames(testes1) <- c("Estatistica", "GL", "P-valor")

testes1

##              Estatistica GL    P-valor
## Razao de Verossimilhanca    3.759049  1 0.05252245
## Wald                      3.339779  1 0.06762367
## Score                      3.759049  1 0.05324632

# Vemos que nenhum dos 3 testes rejeitaram a hipótese nula a 5%. Isto é, o modelo não é significativamente diferente de zero.

coef1 <- summary(fit1)$coefficients

colnames(coef1) <- c("Estimativa", "Erro Padrao", "Valor Z", "Pr(>|z|)")
rownames(coef1) <- c("Intercepto", "Nível de \n fosfatase ácida")

coef1

##              Estimativa Erro Padrao  Valor Z  Pr(>|z|)
## Intercepto          -1.32608905  0.564939407 -2.347312 0.01890941
## Nível de \n fosfatase ácida  0.01400261  0.007662138  1.827506 0.06762367

odds.ratio(fit1)[2,]

## Waiting for profiling to be done...

##              OR    2.5 % 97.5 %      p
## nivel_fosfatase_acida 1.01410 0.99985  1.031 0.06762 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# A Odds-ratio mostra que para cada 1 (x100) unidades do nível de fosfatase ácida, aumenta em 1,41% a chance de ocorrência de doença.
# Entretanto, o intervalo contém o valor 1, o que indica que é pouco significativo este resultado, e pode ser devido ao pequeno tamanho da amostra.

# Medidas de qualidade de ajuste

medidas1 <- as.data.frame(cbind(fit1$deviance, fit1$aic, BIC(fit1),
                               logLik(fit1)[1]))
colnames(medidas1) <- c("Deviance", "AIC", "BIC", "Log Likelihood")
medidas1

##   Deviance      AIC      BIC Log Likelihood
## 1 134.4498 138.4498 143.6998      -67.2249

# Intervalo de confiança para a odds ratio
fit1_ci <- get_confint(fit = fit1,
                      variavel = "nivel_fosfatase_acida",
                      digitos = 5)

## Waiting for profiling to be done...
## Waiting for profiling to be done...

# Plotando com I.C. 95%
temp.data <- data.frame(df$nivel_fosfatase_acida)
colnames(temp.data) = "nivel_fosfatase_acida"

```

```

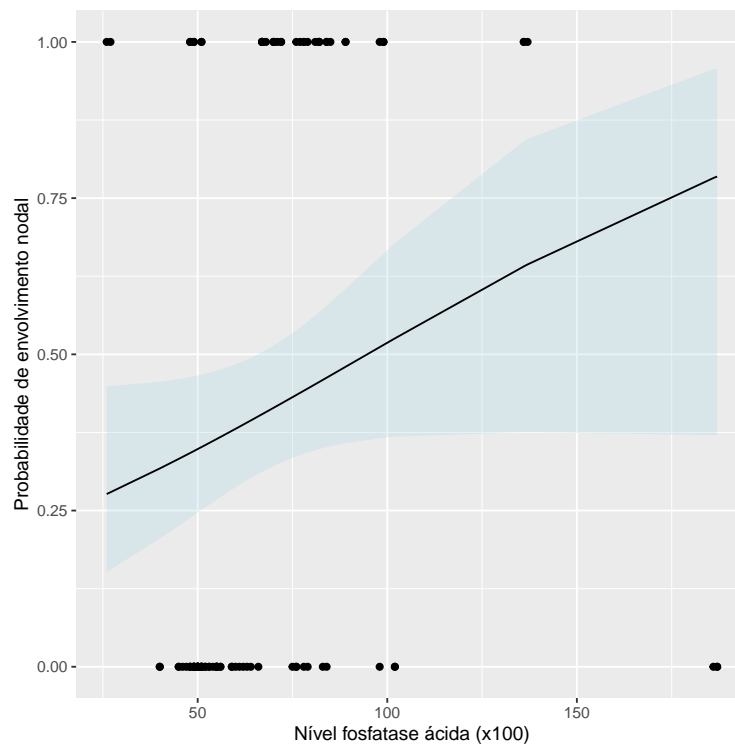
predicted.data <- as.data.frame(predict(fit1, newdata = temp.data,
                                       type="link", se=TRUE))

new.data <- cbind(temp.data, predicted.data)

new.data$yci <- fit1$family$linkinv(new.data$fit - qnorm(0.975) * new.data$se)
new.data$ycs <- fit1$family$linkinv(new.data$fit + qnorm(0.975) * new.data$se)
new.data$fit <- fit1$family$linkinv(new.data$fit)

p <- ggplot(df, aes(x=nivel_fosfatase_acida, y=envolvimento_nodal))
p + geom_point(na.rm = T) +
  geom_ribbon(data=new.data, aes(y=fit, ymin=yci, ymax=ycs),
            fill="lightblue", alpha=0.3) +
  geom_line(data=new.data, aes(y=fit)) +
  labs(x="Nível fosfatase ácida (x100)", y="Probabilidade de envolvimento nodal")

```



```

# Parte 3) Adicionando outras variáveis no modelo ----

# Modelo saturado:
fit2 <- glm(envolvimento_nodal ~ resultado_radiografia +
            + estagio_tumor +
            + nivel_fosfatase_acida,
            family=binomial(link=logit),
            data=df)

summary(fit2); confint(fit2)

##
## Call:

```

```
## glm(formula = envolvimento_nodal ~ resultado_radiografia + +estagio_tumor +
##      nivel_fosfatase_acida, family = binomial(link = logit), data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -5.10392    1.14543  -4.456 8.35e-06 ***
## resultado_radiografia  2.74263    0.73600   3.726 0.000194 ***
## estagio_tumor        3.20837    0.75428   4.254 2.10e-05 ***
## nivel_fosfatase_acida  0.02781    0.00904   3.076 0.002096 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 138.209  on 101  degrees of freedom
## Residual deviance:  86.194  on  98  degrees of freedom
## AIC: 94.194
##
## Number of Fisher Scoring iterations: 5

## Waiting for profiling to be done...

##              2.5 %      97.5 %
## (Intercept)      -7.67761731 -3.10175653
## resultado_radiografia  1.44594984  4.41624960
## estagio_tumor        1.89522148  4.92529003
## nivel_fosfatase_acida  0.01058845  0.04687776

# Com o modelo saturado, todas as variáveis se tornam significativas a 5%, inclusive o nível de fosfatase_acida

hoslem_fit2 <- ResourceSelection::hoslem.test(fit2$y,fit2$fitted.values)
# H_0) Valores observados e valores esperados são iguais para diferentes níveis de nivel_fosfatase_acida
# H_1) c.c.
# O teste de H-L rejeita a hipótese nula a 5%, portanto indica que o modelo não é adequado.

# Outra opção de teste
# DescTools::HosmerLemeshowTest(fit = fitted(fit2),
#                               obs = df$envolvimento_nodal)

# Analisando a influência de cada valor no resultado da regressão logística (análise de resíduos)
stats::influence.measures(fit2)

## Influence measures of
##      glm(formula = envolvimento_nodal ~ resultado_radiografia + +estagio_tumor +      nivel_fosfatase_acida,
##
##              dfb.1_ dfb.rsl_ dfb.est_ dfb.nv__ dffit cov.r  cook.d  hat inf
## 1  -0.03056    0.0217  0.02950  0.01936 -0.0327 1.056 1.20e-04 0.0162
## 2  -0.02761    0.0191  0.02603  0.01841 -0.0290 1.055 9.45e-05 0.0150
## 3  -0.02858    0.0199  0.02715  0.01875 -0.0302 1.056 1.02e-04 0.0154
## 4  -0.02761    0.0191  0.02603  0.01841 -0.0290 1.055 9.45e-05 0.0150
## 5  -0.02714    0.0187  0.02549  0.01823 -0.0284 1.055 9.07e-05 0.0148
## 6  -0.14083   -0.0754  0.13048  0.12218 -0.2528 1.101 8.18e-03 0.0815
## 7  -0.12427   -0.1226  0.13625  0.08786 -0.3055 1.086 1.28e-02 0.0815
## 8  -0.13483   -0.1003  0.13550  0.10697 -0.2818 1.094 1.05e-02 0.0816
```

## 9	0.16310	0.2017	-0.19418	-0.10080	0.4657	1.027	3.97e-02	0.0815	
## 10	-0.02621	0.0179	0.02444	0.01787	-0.0273	1.055	8.37e-05	0.0145	
## 11	-0.02714	0.0187	0.02549	0.01823	-0.0284	1.055	9.07e-05	0.0148	
## 12	-0.02761	0.0191	0.02603	0.01841	-0.0290	1.055	9.45e-05	0.0150	
## 13	-0.04227	0.0344	0.04566	0.01967	-0.0514	1.060	3.01e-04	0.0217	
## 14	-0.04489	0.0380	0.05014	0.01849	-0.0572	1.061	3.74e-04	0.0234	
## 15	-0.05153	0.0506	0.06533	0.00983	-0.0804	1.065	7.49e-04	0.0304	
## 16	-0.02858	0.0199	0.02715	0.01875	-0.0302	1.056	1.02e-04	0.0154	
## 17	0.27652	-0.2747	-0.35424	-0.04741	0.4391	0.845	8.66e-02	0.0310	*
## 18	0.23267	0.1210	0.09557	-0.61610	-0.8098	1.234	1.06e-01	0.2284	*
## 19	-0.09681	0.1368	0.01863	0.15115	0.2313	1.120	6.60e-03	0.0904	
## 20	-0.09770	0.0650	0.00533	0.10216	-0.1775	1.040	4.21e-03	0.0344	
## 21	-0.07453	0.0612	-0.01679	0.07225	-0.1840	1.026	4.75e-03	0.0296	
## 22	-0.07453	0.0612	-0.01679	0.07225	-0.1840	1.026	4.75e-03	0.0296	
## 23	-0.05876	0.0577	-0.03030	0.05232	-0.1886	1.019	5.13e-03	0.0280	
## 24	-0.04411	0.0541	-0.04226	0.03396	-0.1937	1.013	5.55e-03	0.0271	
## 25	-0.04229	0.0893	0.06296	0.01421	0.1013	1.077	1.20e-03	0.0422	
## 26	0.11425	-0.0912	0.02110	-0.11202	0.2706	0.976	1.36e-02	0.0300	
## 27	-0.08004	0.0622	-0.01184	0.07928	-0.1825	1.029	4.62e-03	0.0305	
## 28	0.20256	-0.0228	-0.21270	-0.26673	-0.3751	0.961	3.06e-02	0.0436	
## 29	0.03728	0.0313	-0.10313	-0.06652	-0.2373	0.992	9.44e-03	0.0286	
## 30	-0.03303	0.0433	0.03544	0.02408	0.0487	1.064	2.68e-04	0.0246	
## 31	-0.03553	0.0496	0.03970	0.02441	0.0553	1.066	3.48e-04	0.0269	
## 32	-0.00504	-0.0393	0.07705	0.02618	0.2092	1.004	6.80e-03	0.0271	
## 33	0.00973	-0.0460	0.07070	0.00923	0.2142	1.000	7.29e-03	0.0267	
## 34	-0.05460	-0.0154	0.09626	0.08250	0.1981	1.023	5.64e-03	0.0311	
## 35	0.00973	-0.0460	0.07070	0.00923	0.2142	1.000	7.29e-03	0.0267	
## 36	-0.03784	0.0567	0.04432	0.02409	0.0630	1.068	4.52e-04	0.0295	
## 37	-0.03089	0.0386	0.03216	0.02345	0.0438	1.063	2.17e-04	0.0229	
## 38	-0.02714	0.0187	0.02549	0.01823	-0.0284	1.055	9.07e-05	0.0148	
## 39	-0.03006	0.0213	0.02890	0.01922	-0.0321	1.056	1.16e-04	0.0160	
## 40	-0.02907	0.0204	0.02772	0.01891	-0.0308	1.056	1.07e-04	0.0156	
## 41	-0.02667	0.0183	0.02496	0.01805	-0.0279	1.055	8.71e-05	0.0147	
## 42	-0.14103	-0.0729	0.12968	0.12327	-0.2497	1.102	7.96e-03	0.0815	
## 43	-0.12230	-0.1260	0.13610	0.08457	-0.3090	1.085	1.31e-02	0.0815	
## 44	-0.13590	-0.0973	0.13515	0.10916	-0.2785	1.095	1.02e-02	0.0816	
## 45	-0.03313	0.0242	0.03268	0.01996	-0.0362	1.057	1.48e-04	0.0172	
## 46	0.15618	0.2027	-0.18956	-0.09310	0.4615	1.029	3.87e-02	0.0815	
## 47	-0.02761	0.0191	0.02603	0.01841	-0.0290	1.055	9.45e-05	0.0150	
## 48	-0.02809	0.0195	0.02659	0.01858	-0.0296	1.056	9.84e-05	0.0152	
## 49	-0.04280	0.0351	0.04653	0.01948	-0.0525	1.060	3.14e-04	0.0220	
## 50	-0.04541	0.0388	0.05107	0.01817	-0.0585	1.061	3.91e-04	0.0237	
## 51	-0.02809	0.0195	0.02659	0.01858	-0.0296	1.056	9.84e-05	0.0152	
## 52	-0.04119	0.0330	0.04395	0.01996	-0.0493	1.059	2.76e-04	0.0210	
## 53	0.27999	-0.2749	-0.35499	-0.05341	0.4369	0.842	8.72e-02	0.0304	*
## 54	0.22322	0.1218	0.09801	-0.60034	-0.7933	1.232	1.01e-01	0.2254	*
## 55	-0.09704	0.1348	0.01956	0.15066	0.2282	1.121	6.40e-03	0.0902	
## 56	-0.05108	-0.0172	0.09504	0.07854	0.1987	1.021	5.71e-03	0.0307	
## 57	-0.07160	0.0606	-0.01936	0.06853	-0.1848	1.025	4.82e-03	0.0293	
## 58	-0.05527	0.0569	-0.03320	0.04793	-0.1897	1.018	5.23e-03	0.0277	
## 59	-0.04017	0.0531	-0.04540	0.02904	-0.1952	1.012	5.67e-03	0.0270	
## 60	0.12066	-0.0938	0.01785	-0.11951	0.2751	0.974	1.42e-02	0.0305	
## 61	-0.07734	0.0617	-0.01428	0.07583	-0.1832	1.027	4.69e-03	0.0300	
## 62	0.25958	-0.4345	-0.32752	-0.14298	-0.4825	0.825	1.24e-01	0.0335	*

```

## 63 -0.03260 0.0423 0.03476 0.02398 0.0476 1.064 2.57e-04 0.0243
## 64 -0.03593 0.0507 0.04045 0.02440 0.0565 1.066 3.63e-04 0.0273
## 65 -0.00976 -0.0371 0.07903 0.03158 0.2078 1.006 6.66e-03 0.0273
## 66 -0.03512 0.0485 0.03897 0.02439 0.0541 1.065 3.33e-04 0.0265
## 67 -0.00504 -0.0393 0.07705 0.02618 0.2092 1.004 6.80e-03 0.0271
## 68 0.00973 -0.0460 0.07070 0.00923 0.2142 1.000 7.29e-03 0.0267
## 69 -0.05460 -0.0154 0.09626 0.08250 0.1981 1.023 5.64e-03 0.0311
## 70 0.00973 -0.0460 0.07070 0.00923 0.2142 1.000 7.29e-03 0.0267
## 71 -0.02899 0.1436 0.08595 -0.03069 0.1836 1.100 4.09e-03 0.0700
## 72 -0.04065 0.0324 0.04312 0.02008 -0.0482 1.059 2.64e-04 0.0207
## 73 0.27652 -0.2747 -0.35424 -0.04741 0.4391 0.845 8.66e-02 0.0310 *
## 74 0.23267 0.1210 0.09557 -0.61610 -0.8098 1.234 1.06e-01 0.2284 *
## 75 -0.09681 0.1368 0.01863 0.15115 0.2313 1.120 6.60e-03 0.0904
## 76 -0.05460 -0.0154 0.09626 0.08250 0.1981 1.023 5.64e-03 0.0311
## 77 -0.07453 0.0612 -0.01679 0.07225 -0.1840 1.026 4.75e-03 0.0296
## 78 -0.07453 0.0612 -0.01679 0.07225 -0.1840 1.026 4.75e-03 0.0296
## 79 -0.09770 0.0650 0.00533 0.10216 -0.1775 1.040 4.21e-03 0.0344
## 80 -0.05876 0.0577 -0.03030 0.05232 -0.1886 1.019 5.13e-03 0.0280
## 81 -0.04411 0.0541 -0.04226 0.03396 -0.1937 1.013 5.55e-03 0.0271
## 82 -0.04209 0.0950 0.06586 0.01112 0.1088 1.079 1.38e-03 0.0448
## 83 -0.04229 0.0893 0.06296 0.01421 0.1013 1.077 1.20e-03 0.0422
## 84 0.11425 -0.0912 0.02110 -0.11202 0.2706 0.976 1.36e-02 0.0300
## 85 -0.08004 0.0622 -0.01184 0.07928 -0.1825 1.029 4.62e-03 0.0305
## 86 0.20256 -0.0228 -0.21270 -0.26673 -0.3751 0.961 3.06e-02 0.0436
## 87 -0.01420 0.0462 -0.06550 -0.00320 -0.2067 1.004 6.63e-03 0.0267
## 88 -0.03303 0.0433 0.03544 0.02408 0.0487 1.064 2.68e-04 0.0246
## 89 -0.03632 0.0519 0.04120 0.02438 0.0577 1.066 3.79e-04 0.0277
## 90 -0.00504 -0.0393 0.07705 0.02618 0.2092 1.004 6.80e-03 0.0271
## 91 -0.03553 0.0496 0.03970 0.02441 0.0553 1.066 3.48e-04 0.0269
## 92 -0.05460 -0.0154 0.09626 0.08250 0.1981 1.023 5.64e-03 0.0311
## 93 0.00973 -0.0460 0.07070 0.00923 0.2142 1.000 7.29e-03 0.0267
## 94 -0.03784 0.0567 0.04432 0.02409 0.0630 1.068 4.52e-04 0.0295
## 95 -0.03089 0.0386 0.03216 0.02345 0.0438 1.063 2.17e-04 0.0229
## 96 -0.02767 0.1463 0.08682 -0.03383 0.1884 1.101 4.32e-03 0.0716
## 97 -0.02714 0.0187 0.02549 0.01823 -0.0284 1.055 9.07e-05 0.0148
## 98 -0.03006 0.0213 0.02890 0.01922 -0.0321 1.056 1.16e-04 0.0160
## 99 -0.02809 0.0195 0.02659 0.01858 -0.0296 1.056 9.84e-05 0.0152
## 100 -0.02907 0.0204 0.02772 0.01891 -0.0308 1.056 1.07e-04 0.0156
## 101 -0.02809 0.0195 0.02659 0.01858 -0.0296 1.056 9.84e-05 0.0152
## 102 -0.14103 -0.0729 0.12968 0.12327 -0.2497 1.102 7.96e-03 0.0815

# teste verossimilhanca
t2 <- lr.test(fit0,fit2)
# O teste de razão de verossimilhança rejeita a hipótese nula, portanto o modelo saturado é preferível

# teste score
score2<-anova(fit0,fit2, test="Rao")
resultados <- cbind(t2$LR,3,t2$pvalue)
resultados3 <- cbind(anova(fit0,fit2, test="Rao")[2,4],
                     anova(fit0,fit2, test="Rao")[2,3],
                     anova(fit0,fit2, test="Rao")[2,6])

testes2 <- rbind(resultados,resultados3)
rownames(testes2) <- c("Razao de Verossimilhanca","Score")

```



```

colnames(testes2) <- c("Estatistica","GL","P-valor")

testes2

##              Estatistica GL      P-valor
## Razao de Verossimilhanca    52.01516  3 2.972887e-11
## Score                      52.01516  3 3.680336e-09

# Vemos que ambos os testes rejeitaram a hipótese nula a 5%.
# Isto é, o modelo é significativamente melhor que o modelo que contém apenas o intercepto.

coef2 <- summary(fit2)$coefficients

colnames(coef2) <- c("Estimativa","Erro Padrao","Valor Z","Pr(>|z|)")
rownames(coef2) <- c("Intercepto",
                    "Resultado da Radiografia",
                    "Estágio do Tumor",
                    "Nível de Fosfatase Ácida"
                    )

coef2

##              Estimativa Erro Padrao  Valor Z    Pr(>|z|)
## Intercepto          -5.10391894  1.145431071 -4.455894 8.354447e-06
## Resultado da Radiografia  2.74263064  0.735996119  3.726420 1.942187e-04
## Estágio do Tumor          3.20836574  0.754274940  4.253576 2.103835e-05
## Nível de Fosfatase Ácida  0.02780885  0.009039928  3.076225 2.096398e-03

odds.ratio(fit2)[2,]

## Waiting for profiling to be done...

##              OR    2.5 % 97.5 %          p
## resultado_radiografia 15.5278  4.2459 82.785 0.0001942 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# A odds ratio mostra que o resultado positivo na radiografia aumenta de 4,246 a 82,785 a chance de haver

odds.ratio(fit2)[3,]

## Waiting for profiling to be done...

##              OR    2.5 % 97.5 %          p
## estagio_tumor 24.739  6.654 137.73 2.104e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# A odds ratio mostra que o fato do tumor ser mais grave aumenta de 6,654 a 137,73 a chance de haver en

odds.ratio(fit2)[4,]

## Waiting for profiling to be done...

##              OR    2.5 % 97.5 %          p
## nivel_fosfatase_acida 1.0282 1.0106  1.048 0.002096 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

A Odds-ratio mostra que para cada 1 (x100) unidades do nível de fosfatase ácida, aumenta em 2,82% a chance de ser positivo para o câncer
Diferente do modelo simples, o intervalo não contém o valor 1, indicando que é positiva a relação dessa variável com o resultado da radiografia

```
fit2_odds <-
odds.ratio(fit2) |>
  as_tibble() |>
  slice(-1) |>
  mutate(Variável = c("Resultado da Radiografia",
                      "Estágio do Tumor",
                      "Nível de Fosfatase Ácida")) |>
dplyr::select(Variável, everything()) |>
  rename("Estimativa Pontual" = OR, "P-valor" = p)

## Waiting for profiling to be done...

# Essas duas variáveis são extremamente significativas, e contribuem fortemente para a explicação de ha

# Medidas de qualidade de ajuste
medidas2 <- as.data.frame(cbind(fit2$deviance, fit2$AIC, BIC(fit2),
                              logLik(fit2)[1]))
colnames(medidas2) <- c("Deviance", "AIC", "BIC", "Log Likelihood")
medidas2

##      Deviance      AIC      BIC Log Likelihood
## 1  86.1937  94.1937 104.6936      -43.09685

# Outros modelos possíveis:
fit3 <- glm(envolvimento_nodal ~ resultado_radiografia + estagio_tumor,
            family=binomial(link=logit),
            data=df)

medidas3 <- as.data.frame(cbind(fit3$deviance, fit3$AIC, BIC(fit3),
                              logLik(fit3)[1]))
colnames(medidas3) <- c("Deviance", "AIC", "BIC", "Log Likelihood")

fit4 <- glm(envolvimento_nodal ~ resultado_radiografia + nivel_fosfatase_acida,
            family=binomial(link=logit),
            data=df)

medidas4 <- as.data.frame(cbind(fit4$deviance, fit4$AIC, BIC(fit4),
                              logLik(fit4)[1]))
colnames(medidas4) <- c("Deviance", "AIC", "BIC", "Log Likelihood")

fit5 <- glm(envolvimento_nodal ~ nivel_fosfatase_acida + estagio_tumor,
            family=binomial(link=logit),
            data=df)

medidas5 <- as.data.frame(cbind(fit5$deviance, fit5$AIC, BIC(fit5),
                              logLik(fit5)[1]))
colnames(medidas5) <- c("Deviance", "AIC", "BIC", "Log Likelihood")

fit6 <- glm(envolvimento_nodal ~ estagio_tumor,
            family=binomial(link=logit),
            data=df)
```

```

medidas6 <- as.data.frame(cbind(fit6$deviance, fit6$aic, BIC(fit6),
                                logLik(fit6)[1]))
colnames(medidas6) <- c("Deviance", "AIC", "BIC", "Log Likelihood")

fit7 <- glm(envolvimento_nodal ~ resultado_radiografia,
            family=binomial(link=logit),
            data=df)

medidas7 <- as.data.frame(cbind(fit7$deviance, fit7$aic, BIC(fit7),
                                logLik(fit7)[1]))
colnames(medidas7) <- c("Deviance", "AIC", "BIC", "Log Likelihood")

# Comparando os modelos

fit = glm(
  envolvimento_nodal ~ nivel_fosfatase_acida + resultado_radiografia + estagio_tumor,
  family = binomial(link = logit),
  data = df
)

car::Anova(fit) # A interação não é significativa.

## Analysis of Deviance Table (Type II tests)
##
## Response: envolvimento_nodal
##
##          LR Chisq Df Pr(>Chisq)
## nivel_fosfatase_acida  9.8889  1  0.001663 **
## resultado_radiografia 20.4490  1  6.124e-06 ***
## estagio_tumor        30.3447  1  3.617e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

stepAIC(fit, direction = c("both")) # Voltamos ao modelo saturado.

## Start:  AIC=94.19
## envolvimento_nodal ~ nivel_fosfatase_acida + resultado_radiografia +
##      estagio_tumor
##
##          Df Deviance      AIC
## <none>          86.194  94.194
## - nivel_fosfatase_acida  1   96.083 102.083
## - resultado_radiografia  1  106.643 112.643
## - estagio_tumor         1  116.538 122.538
##
## Call:  glm(formula = envolvimento_nodal ~ nivel_fosfatase_acida + resultado_radiografia +
##      estagio_tumor, family = binomial(link = logit), data = df)
##
## Coefficients:
##      (Intercept) nivel_fosfatase_acida resultado_radiografia
##          -5.10392           0.02781           2.74263
##      estagio_tumor
##          3.20837
##

```

```
## Degrees of Freedom: 101 Total (i.e. Null); 98 Residual
## Null Deviance: 138.2
## Residual Deviance: 86.19 AIC: 94.19

rm(fit)

stepwise <- stepAIC(fit2, direction = c("both"))

## Start: AIC=94.19
## envolvimento_nodal ~ resultado_radiografia + estagio_tumor +
## nivel_fosfatase_acida
##
##
## Df Deviance AIC
## <none> 86.194 94.194
## - nivel_fosfatase_acida 1 96.083 102.083
## - resultado_radiografia 1 106.643 112.643
## - estagio_tumor 1 116.538 122.538
```

O método stepwise indica ficar com o modelo saturado.

```
Modelo = c("X4 ~ 1", "X4 ~ X3", "X4 ~ X1 + X2 + X3", "X4 ~ X1 + X2", "X4 ~ X1 + X3", "X4 ~ X2 + X3", "X4 ~ X2 + X3", "X4 ~ X2 + X3")
medidas = rbind(medidas0, medidas1, medidas2, medidas3, medidas4, medidas5, medidas6, medidas7)
medidas$Modelo = Modelo
medidas = medidas[c(1,3,4,6,8,7,2), c(5,1:4)]
```

```
kable(medidas)
```

	Modelo	Deviance	AIC	BIC	Log Likelihood
1	X4 ~ 1	138.20886	140.2089	142.8338	-69.10443
3	X4 ~ X1 + X2 + X3	86.19370	94.1937	104.6936	-43.09685
4	X4 ~ X1 + X2	96.08257	102.0826	109.9575	-48.04129
5	X4 ~ X1 + X3	116.53844	122.5384	130.4134	-58.26922
6	X4 ~ X2 + X3	106.64265	112.6427	120.5176	-53.32133
8	X4 ~ X1	119.87619	123.8762	129.1261	-59.93810
7	X4 ~ X2	116.02919	120.0292	125.2791	-58.01460
2	X4 ~ X3	134.44981	138.4498	143.6998	-67.22490

Pela tabela, devemos optar pelo modelo saturado, pois:
1) tem menor deviance 2) tem menor AIC 3) tem maior BIC 4) tem maior log-verossimilhança.

```
# rm(fit0, fit1, fit3, fit4, fit5, fit6, fit7,
# medidas0, medidas1, medidas3, medidas4, medidas5, medidas6, medidas7,
# resultados, resultados2, resultados3, p, predicted.data, vcov1, Modelo, thetahat,
# WaldTest, LL, new.data, score1, t1, testes1, temp.data, coef1, t2)
```

Métricas do modelo escolhido ----

```
kable(medidas2)
```

Deviance	AIC	BIC	Log Likelihood
86.1937	94.1937	104.6936	-43.09685

```
kable(coef2)
```

	Estimativa	Erro Padrao	Valor Z	Pr(> z)
Intercepto	-5.1039189	1.1454311	-4.455894	0.0000084
Resultado da Radiografia	2.7426306	0.7359961	3.726420	0.0001942
Estágio do Tumor	3.2083657	0.7542749	4.253576	0.0000210
Nível de Fosfatase Ácida	0.0278089	0.0090399	3.076225	0.0020964

```
# Curva ROC do modelo fit2
```

```
ROC <- roc(response = df$envolvimento_nodal, predictor = predict(fit2, type = "response"))
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
# Create a data frame from the ROC object for plotting
```

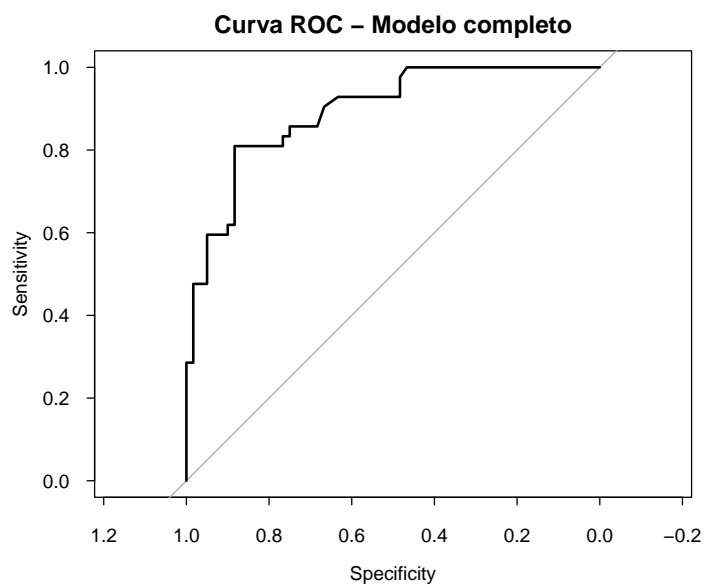
```
roc_data <- data.frame(
```

```
  Spec_comp = 1 - ROC$specificities, # False Positive Rate = 1 - Specificity
```

```
  Sensit = ROC$sensitivities          # True Positive Rate = Sensitivity
```

```
)
```

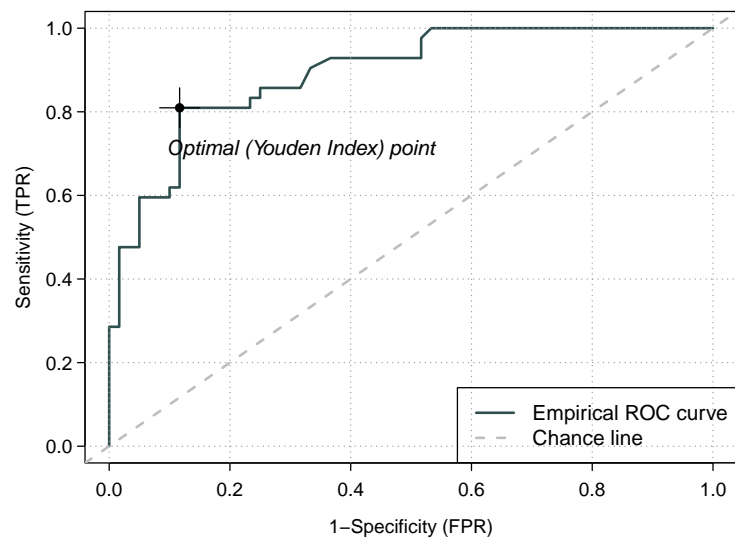
```
plot(ROC, main = "Curva ROC - Modelo completo")
```



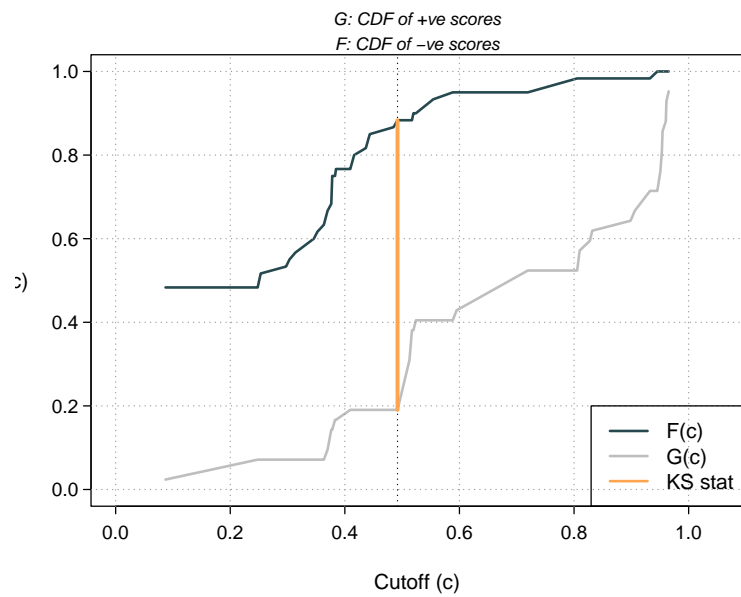
```
# Alternativo:
```

```
ROC2 <- rocit(score=predict(fit2, type = "response"), class=df$envolvimento_nodal)
```

```
plot(ROC2)
```



```
ksplot(ROC2)
```



```
performance_hosmer(fit2, n_bins = 10)
```

```
## # Hosmer-Lemeshow Goodness-of-Fit Test
```

```
##
##      Chi-squared: 22.054
##              df: 8
##      p-value: 0.005

## Summary: model does not fit well.

# Aplicando o modelo aos dados de teste:
teste$predict <- predict(fit2, teste,
                        type = "response")

# Usando um threshold = 0.5...
teste = teste %>%
  mutate(env_pred = ifelse(predict > .5, 1, 0))

# Matriz de confusão:
table_confusao <- table(teste$envolvimento_nodal, teste$env_pred)
kable(table_confusao)
```

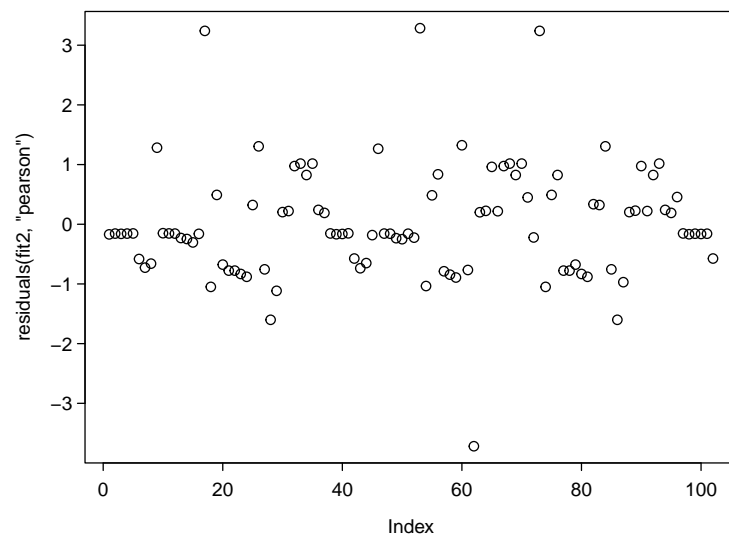
	0	1
0	54	10
1	10	28

```
acertos <- sum(diag(table_confusao)) / sum(table_confusao)

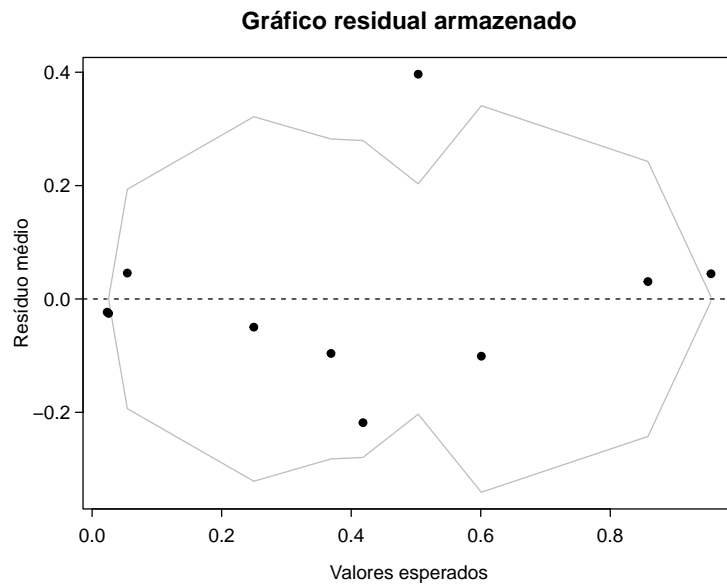
matriz_conf_alternativa <- table(teste$envolvimento_nodal, teste$env_pred) |>
  as.data.frame() |>
  pivot_wider(values_from = Freq,
              names_from = Var2) |>
  rename("Observado\\Predito" = Var1)

# Para os dados de validação, o modelo teve 54+28/102 = 80,4% de acerto, o que é bastante razoável.

# Análise de resíduos
plot(residuals(fit2, "pearson"))
```

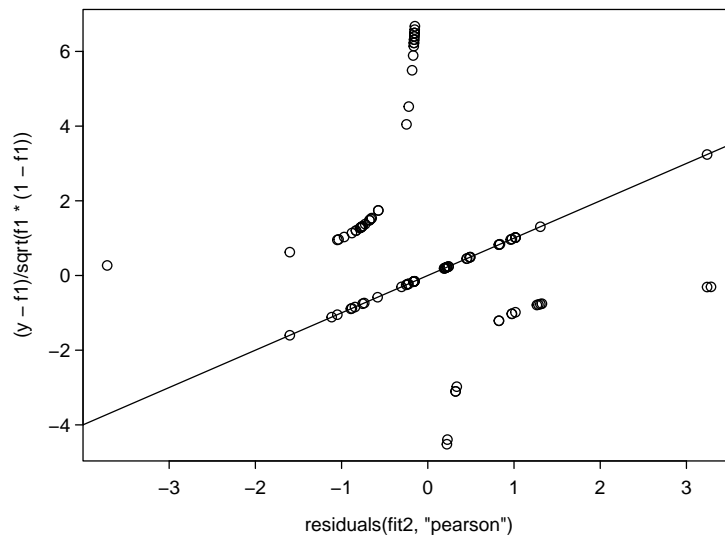


```
# Outra visualização: Binned residual plot
binnedplot(fitted(fit2),
  residuals(fit2, type = "response"),
  nclass = NULL,
  xlab = "Valores esperados",
  ylab = "Resíduo médio",
  main = "Gráfico residual armazenado",
  cex.pts = 0.8,
  col.pts = 1,
  col.int = "gray")
```

```
#Deveríamos observar 95% das observações sobre o intervalo produzido pelo ASE (linhas cinzas)

# Outra que achei na internet (acho que não faz muito sentido na real)
y <- rbinom(102, 1, 0.5)
f1 <- fitted(fit2)
plot( residuals(fit2, "pearson"), (y-f1)/sqrt(f1*(1-f1)))
abline(0,1)
```



```
# Worm plot (está se popularizando para análise de MLGs)
p_load(gamlss)

## Installing package into '/home/rafaelrocha/R/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
## also installing the dependencies 'gamlss.data', 'gamlss.dist'
##
## gamlss installed

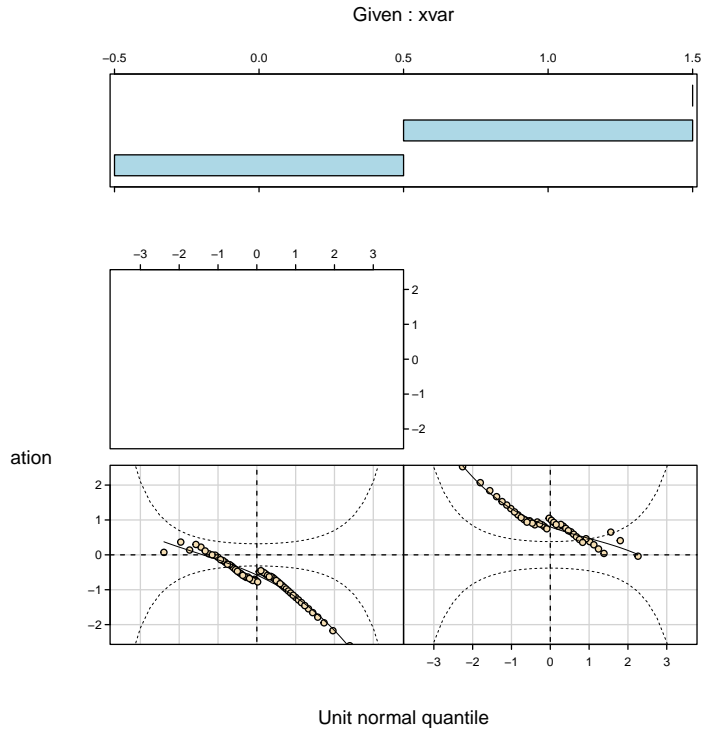
wp(resid=resid(fit2), xvar=df$envolvimento_nodal)

## number of missing points from plot= 1 out of 60

## Warning in panel(x[id], y[id], col = col[id], pch = pch[id], ...): Some points are missed
out
## increase the y limits using ylim.worm

## number of missing points from plot= 1 out of 42

## Warning in panel(x[id], y[id], col = col[id], pch = pch[id], ...): Some points are missed
out
## increase the y limits using ylim.worm
```



*# Pelo worm plot, o ajuste do modelo parece muito ruim.
 # Mas isso aparenta ter mais a ver com a técnica (regressão logística) do que quanto ao modelo
 # específico ajustado.*

The R session information (including the OS info, R version and all packages used):

```
sessionInfo()

## R version 4.4.0 (2024-04-24)
## Platform: x86_64-pc-linux-gnu
## Running under: Debian GNU/Linux 11 (bullseye)
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblas-r0.3.13.so; LAPACK version 3.9.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=pt_BR.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=pt_BR.UTF-8  LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=pt_BR.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=pt_BR.UTF-8 LC_IDENTIFICATION=C
##
## time zone: America/Sao_Paulo
## tzcode source: system (glibc)
##
## attached base packages:
## [1] parallel splines stats      graphics grDevices utils      datasets
## [8] methods  base
```

```
##
## other attached packages:
## [1] gamlss_5.4-22          nlme_3.1-152          gamlss.dist_6.1-1
## [4] gamlss.data_6.0-6      pacman_0.5.1          DescTools_0.99.54
## [7] arm_1.14-4            lme4_1.1-35.5        Matrix_1.7-0
## [10] performance_0.12.0     compareGroups_4.8.0   labelled_2.13.0
## [13] ROCit_2.1.2           pROC_1.18.5          knitr_1.47
## [16] car_3.1-2             carData_3.0-5         lmtest_0.9-40
## [19] zoo_1.8-12            ResourceSelection_0.3-6 mlpack_4.4.0
## [22] questionr_0.7.8       AICcmmodavg_2.3-3     mdscore_0.1-3
## [25] MASS_7.3-53.1         cowplot_1.1.3         lubridate_1.9.3
## [28] forcats_1.0.0         stringr_1.5.1         dplyr_1.1.4
## [31] purrr_1.0.2           readr_2.1.5           tidyr_1.3.1
## [34] tibble_3.2.1          ggplot2_3.5.1         tidyverse_2.0.0
## [37] readxl_1.4.3
##
## loaded via a namespace (and not attached):
## [1] later_1.3.2           HardyWeinberg_1.7.8   cellranger_1.1.0
## [4] datawizard_0.12.0     rpart_4.1-15          lifecycle_1.0.4
## [7] Rsolnp_1.16           lattice_0.20-41       insight_0.20.1
## [10] backports_1.5.0       magrittr_2.0.3        rmarkdown_2.27
## [13] yaml_2.3.8            httpuv_1.6.15         zip_2.3.1
## [16] askpass_1.2.0         gld_2.6.6            minqa_1.2.7
## [19] abind_1.4-5           expm_0.999-9          nnet_7.3-15
## [22] gdtools_0.3.7         crul_1.4.2            svglite_2.1.3
## [25] codetools_0.2-18     xml2_1.3.6            tidysselect_1.2.1
## [28] shape_1.4.6.1         ggeffects_1.7.0       httpcode_0.3.0
## [31] farver_2.1.2          stats4_4.4.0          jsonlite_1.8.8
## [34] e1071_1.7-14          mitml_0.4-5           survival_3.2-7
## [37] iterators_1.0.14      systemfonts_1.1.0     foreach_1.5.2
## [40] tools_4.4.0           chron_2.3-61          ragg_1.3.2
## [43] Rcpp_1.0.12           glue_1.7.0            pan_1.9
## [46] xfun_0.44             withr_3.0.0           fastmap_1.2.0
## [49] boot_1.3-27           fansi_1.0.6           openssl_2.2.0
## [52] digest_0.6.35         truncnorm_1.0-9       timechange_0.3.0
## [55] R6_2.5.1             mime_0.12             mice_3.16.0
## [58] textshaping_0.4.0     colorspace_2.1-0      utf8_1.2.4
## [61] generics_0.1.3        fontLiberation_0.1.0  data.table_1.15.4
## [64] class_7.3-18          sjPlot_2.8.16         httr_1.4.7
## [67] pkgconfig_2.0.3       gtable_0.3.5          Exact_3.2
## [70] htmltools_0.5.8.1     fontBitstreamVera_0.1.1 scales_1.3.0
## [73] kableExtra_1.4.0      lmom_3.0              rstudioapi_0.16.0
## [76] tzdb_0.4.0           uuid_1.2-0            coda_0.19-4.1
## [79] curl_5.2.1           nloptr_2.1.1          proxy_0.4-27
## [82] flextable_0.9.6       sjlabelled_1.2.0      rootSolve_1.8.2.4
## [85] miniUI_0.1.1.1        pillar_1.9.0          grid_4.4.0
## [88] vctrs_0.6.5          VGAM_1.1-11           promises_1.3.0
## [91] jomo_2.7-6           xtable_1.8-4          evaluate_0.24.0
## [94] mvtnorm_1.2-5         cli_3.6.2             compiler_4.4.0
## [97] rlang_1.1.4          crayon_1.5.2          labeling_0.4.3
## [100] plyr_1.8.9           sjmisc_2.8.10         writexl_1.5.0
## [103] stringi_1.8.4         viridisLite_0.4.2     munsell_0.5.1
## [106] glmnet_4.1-8          fontquiver_0.2.1      sjstats_0.19.0
## [109] hms_1.1.3            unmarked_1.4.1        gfonts_0.2.0
```

```
## [112] shiny_1.8.1.1      highr_0.11      haven_2.5.4
## [115] broom_1.0.6        officer_0.6.6

Sys.time()

## [1] "2024-07-14 16:49:29 -03"
```