The results below are generated from an R script.

```
# Assignment: ASSIGNMENT 9
# Name: Reppeto, Brian
# Date: 2023-07-31

## Set the working directory to the root of your DSC 520 directory
setwd("~/DSC520/Week 9")

library(tidyverse)
library(foreign)

## Load the `data`
surgery_df <- read.arff("ThoracicSurgery.arff")
surgery <- as_tibble(surgery_df)
surgery <- surgery %>%
  rename(
    Diagnosis = DGN,
    FVC = PRE4,
    FEV1 = PRE5,
    Performance = PRE6,
    Pain = PRE7,
    Haemoptysis = PRE8,
    Dyspnoea = PRE9,
    Cough = PRE10,
    Weakness = PRE11,
    Tumor_size = PRE14,
    Diabetes_mellitus = PRE17,
    MI = PRE19,
    PAD = PRE25,
    Smoking = PRE30,
    Asthma = PRE32,
    Risk1Y = Risk1Yr
  )

#surgery$Risk1Y <- factor(surgery$Risk1Y, levels = c("T", "F"))

#survived_patients <- with(surgery, Risk1Y == "T")

model <-
  glm(
    Risk1Y ~ Diagnosis + FVC + FEV1 + Performance + Pain
    + Haemoptysis + Dyspnoea + Cough + Weakness + Tumor_size + Diabetes_mellitus
    + MI + PAD + Smoking + Asthma + AGE,
    data = surgery,
    family = binomial
  )

summary(model)
```

```
##
## Call:
## glm(formula = Risk1Y ~ Diagnosis + FVC + FEV1 + Performance +
##     Pain + Haemoptysis + Dyspnoea + Cough + Weakness + Tumor_size +
##     Diabetes_mellitus + MI + PAD + Smoking + Asthma + AGE, family = binomial,
```

```
##     data = surgery)
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -1.655e+01  2.400e+03  -0.007  0.99450
## DiagnosisDGN2       1.474e+01  2.400e+03   0.006  0.99510
## DiagnosisDGN3       1.418e+01  2.400e+03   0.006  0.99528
## DiagnosisDGN4       1.461e+01  2.400e+03   0.006  0.99514
## DiagnosisDGN5       1.638e+01  2.400e+03   0.007  0.99455
## DiagnosisDGN6       4.089e-01  2.673e+03   0.000  0.99988
## DiagnosisDGN8       1.803e+01  2.400e+03   0.008  0.99400
## FVC                -2.272e-01  1.849e-01  -1.229  0.21909
## FEV1               -3.030e-02  1.786e-02  -1.697  0.08971 .
## PerformancePRZ1    -4.427e-01  5.199e-01  -0.852  0.39448
## PerformancePRZ2    -2.937e-01  7.907e-01  -0.371  0.71030
## PainT               7.153e-01  5.556e-01   1.288  0.19788
## HaemoptysisT        1.743e-01  3.892e-01   0.448  0.65419
## DyspnoeaT           1.368e+00  4.868e-01   2.811  0.00494 **
## CoughT              5.770e-01  4.826e-01   1.196  0.23185
## WeaknessT           5.162e-01  3.965e-01   1.302  0.19295
## Tumor_sizeOC12      4.394e-01  3.301e-01   1.331  0.18318
## Tumor_sizeOC13      1.179e+00  6.165e-01   1.913  0.05580 .
## Tumor_sizeOC14      1.653e+00  6.094e-01   2.713  0.00668 **
## Diabetes_mellitusT  9.266e-01  4.445e-01   2.085  0.03709 *
## MIT                -1.466e+01  1.654e+03  -0.009  0.99293
## PADT               -9.789e-02  1.003e+00  -0.098  0.92227
## SmokingT            1.084e+00  4.990e-01   2.172  0.02984 *
## AsthmaT            -1.398e+01  1.645e+03  -0.008  0.99322
## AGE                -9.506e-03  1.810e-02  -0.525  0.59944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 395.61  on 469  degrees of freedom
## Residual deviance: 341.19  on 445  degrees of freedom
## AIC: 391.19
##
## Number of Fisher Scoring iterations: 15

## DyspnoeaT, Tumor_sizeOC14, Diabetes_mellitusT, and SmokingT all had p-values
## less than 0.05 which indicated that they have the greatest effect on the
## survival rate among the predictor variables included in the model.


# Make predictions using the fitted model
predicted_probs <- predict(model, surgery, type = "response")

# Convert predicted prob to binary predictions (T or F) on a threshold (0.5)
predicted_outcomes <- ifelse(predicted_probs >= 0.5, "T", "F")

# Compare predicted outcomes with actual outcomes and calculate accuracy
correct_predictions <- sum(predicted_outcomes == surgery$Risk1Y)
total_predictions <- nrow(surgery)
```

```r
accuracy <- correct_predictions / total_predictions * 100

# Print the accuracy
cat("Accuracy of the model:", accuracy, "%\n")

## Accuracy of the model: 83.61702 %
```

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

```r
sessionInfo()

## R version 4.3.0 (2023-04-21)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Ventura 13.4.1
##
## Matrix products: default
## BLAS:   /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versio
## LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib;  LAPACK ve
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] foreign_0.8-84  lm.beta_1.7-2   lubridate_1.9.2 forcats_1.0.0   stringr_1.5.0
##  [6] dplyr_1.1.2     purrr_1.0.1     tidyr_1.3.0     tibble_3.2.1    tidyverse_2.0.0
## [11] readxl_1.4.3    caret_6.0-94    lattice_0.21-8  ggplot2_3.4.2   readr_2.1.4
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.2.0     timeDate_4022.108    farver_2.1.1         fastmap_1.1.1
##  [5] pROC_1.18.4          digest_0.6.33        rpart_4.1.19         timechange_0.2.0
##  [9] lifecycle_1.0.3      survival_3.5-5       magrittr_2.0.3       compiler_4.3.0
## [13] rlang_1.1.1          tools_4.3.0          utf8_1.2.3           yaml_2.3.7
## [17] data.table_1.14.8    knitr_1.43           labeling_0.4.2       bit_4.0.5
## [21] plyr_1.8.8           withr_2.5.0          nnet_7.3-19          grid_4.3.0
## [25] stats4_4.3.0         fansi_1.0.4          xtable_1.8-4         colorspace_2.1-0
## [29] future_1.33.0        globals_0.16.2       scales_1.2.1         iterators_1.0.14
## [33] MASS_7.3-60          cli_3.6.1            rmarkdown_2.23       crayon_1.5.2
## [37] generics_0.1.3       rstudioapi_0.15.0    future.apply_1.11.0  reshape2_1.4.4
## [41] tzdb_0.4.0           splines_4.3.0        parallel_4.3.0       cellranger_1.1.0
## [45] vctrs_0.6.3          hardhat_1.3.0        Matrix_1.6-0         hms_1.1.3
## [49] bit64_4.0.5          listenv_0.9.0        foreach_1.5.2        gower_1.0.1
## [53] recipes_1.0.6        glue_1.6.2           parallelly_1.36.0    codetools_0.2-19
## [57] stringi_1.7.12       gtable_0.3.3         munsell_0.5.0        pillar_1.9.0
## [61] htmltools_0.5.5      ipred_0.9-14         lava_1.7.2.1         R6_2.5.1
## [65] vroom_1.6.3          evaluate_0.21        highr_0.10           class_7.3-22
## [69] Rcpp_1.0.11          nlme_3.1-162         prodlim_2023.03.31   xfun_0.39
## [73] ModelMetrics_1.2.2.2 pkgconfig_2.0.3

Sys.time()
```

```
## [1] "2023-08-06 21:36:57 EDT"
```