

```

---
title: "9.2 Logistic Regression"
author: "Moshe Burnstein"
date: "`r Sys.Date()`"
output: pdf_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

# Logistic Regression
## Thoracic Surgery
library(foreign)
thoracic <- read.arff("ThoracicSurgery.arff")

```{r}
library(foreign)
str(thoracic)
head(thoracic)
thoracic_outcome <- glm(Risk1Yr ~ DGN + PRE4 + PRE5 + PRE6 + PRE7 + PRE8 +
PRE9 + PRE10 + PRE11 + PRE14 + PRE17 + PRE19 + PRE25 + PRE30 + PRE32 + AGE,
data = thoracic, family = binomial())
summary(thoracic_outcome)
glm.predict <- predict(thoracic_outcome, thoracic, type = "response")
glm.predict
thoracic$glm.predict <- ifelse(glm.predict >= .5, "T", "F")
thoracic$glm.predict
accuracy_glm.predict <- mean(thoracic$glm.predict == thoracic$Risk1Yr)
accuracy_glm.predict

```

The greatest negative correlation is PRE19T at -14.66, followed by PRE32T at
-13.98. The greatest positive correlations are DGNDGN8 at 18.03 and DGNDGN5 at
16.38, followed by DGNDGN2 at 14.74, DGNDGN4 at 14.61, and DGNDGN3 at 14.18.
The accuracy of this model is 83.6%.

### Binary Classifier GLM

```{r}
binary_df <- read.csv("binary-classifier-data.csv")
binary_df
str(binary_df)

```

```

binary_glm <- glm(label ~ x + y, data = binary_df, family = binomial())
summary(binary_glm)
binary_predict <- predict(binary_glm, binary_df, type = "response")
binary_predict
binary_df$binary_predict <- ifelse(binary_predict >= .5, 1, 0)
head(binary_df$binary_predict)
accuracy_binary_predict <- mean(binary_df$binary_predict == binary_df$label)
accuracy_binary_predict
```

```{r}
```

# The accuracy of this model is 58.3%

```