Example Quarto

Alice Paul

Data

In this report, we use the `NHANESsample` dataset from the HDSinRdata package, which comes from NHANES 1999-2018 and was downloaded from the **nhanesA** package. This data set includes information on lead levels, blood pressure, and demographic variables for 31,265 subjects.

Exploratory Analysis and Model Fitting

We first created a subset of the NHANESsample dataset that contains only the age, sex, lead level, and hypertension status variables. Then, we fit a logistic regression with hypertension as the outcome and include the main effects of all covariates. The coefficients from this model can be seen in Table Table 1.

Table 1: Final Model Intercepts and Odds Ratios

Term	Estimate	Std. Error	P Value
(Intercept)	0.095	0.042	0.000
AGE	1.059	0.001	0.000
SEXFemale	0.663	0.026	0.000
LEAD	0.997	0.007	0.702

To evaluate our model, we use the AUC and ROC curve, shown in Figure 1. The AUC for the model is 0.762. We can further examine the relationship between the predictors in our model and our outcome of interest by drawing our attention to the plots in Figure 2 below.
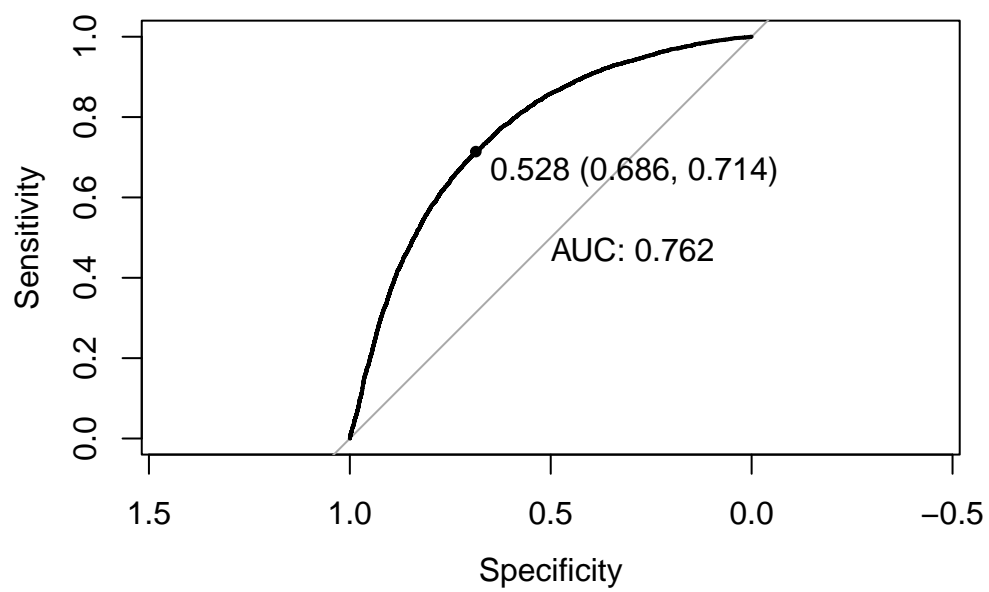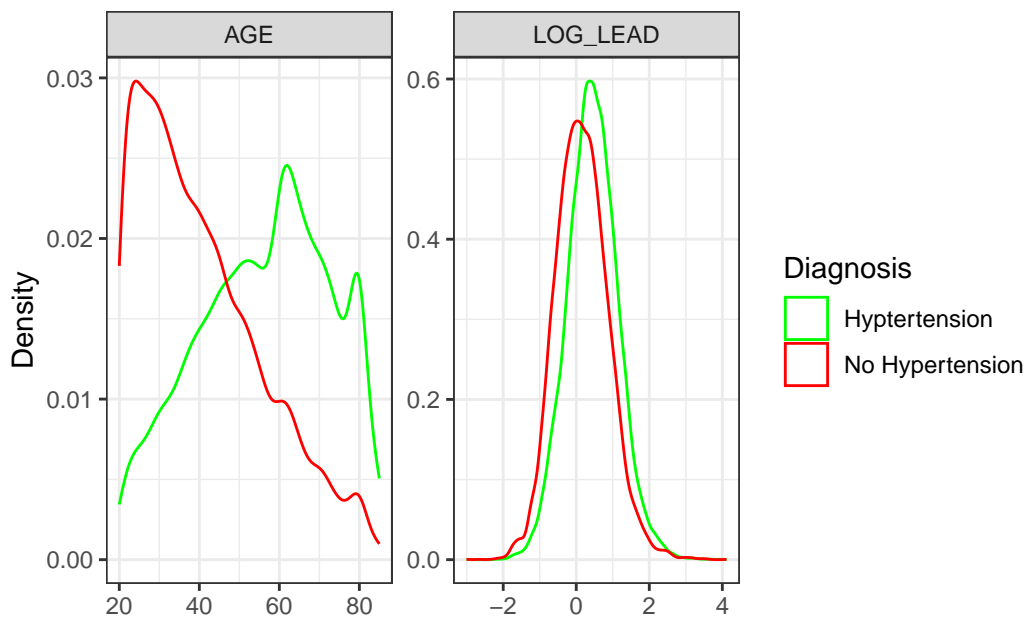
Figure 1: ROC Curve

Figure 2: Predictor Distributions by Outcome

Code Appendix

```r
knitr::opts_chunk$set(message=FALSE,
                      warning=FALSE,
                      error=FALSE,
                      echo = FALSE)
library(tidyverse)
library(knitr)
library(kableExtra)
library(broom)
library(HDSinRdata)
library(pROC)
#load in data
data(NHANESsample)
nhanes <- NHANESsample %>%
  dplyr::select(c(HYP, LEAD, AGE, SEX))

# Fit model
simple_mod <- glm(HYP ~ AGE + SEX + LEAD, data = nhanes,
                  family = binomial)

# Display coefs
tidy(simple_mod, exponentiate=TRUE) %>%
  select(-statistic) %>%
  kable(col.names = c("Term", "Estimate", "Std. Error", "P Value"),
        digits = 3)

# ROC curve
roccurve <- roc(predictor=predict(simple_mod, type="response"),
                response=as.factor(simple_mod$y),
                levels = c(0,1), direction = "<")
plot(roccurve, print.auc=TRUE, print.thres = TRUE)

nhanes_long <- nhanes %>%
  mutate(LOG_LEAD = log(LEAD)) %>%
  select(HYP, AGE, LOG_LEAD) %>%
  pivot_longer(cols = AGE:LOG_LEAD) %>%
  mutate(hypertension_status = case_when(HYP == 1 ~ "Hyptertension",
                                         HYP == 0 ~ "No Hypertension"))

ggplot(data = nhanes_long) +
```

```r
geom_density(aes(x = value, y = ..density.., color = hypertension_status)) +
facet_wrap(~ name, scales = "free") +
labs(y = "Density", x = "") +
scale_color_manual(name = "Diagnosis", values = c("green", "red")) +
theme_bw()
```