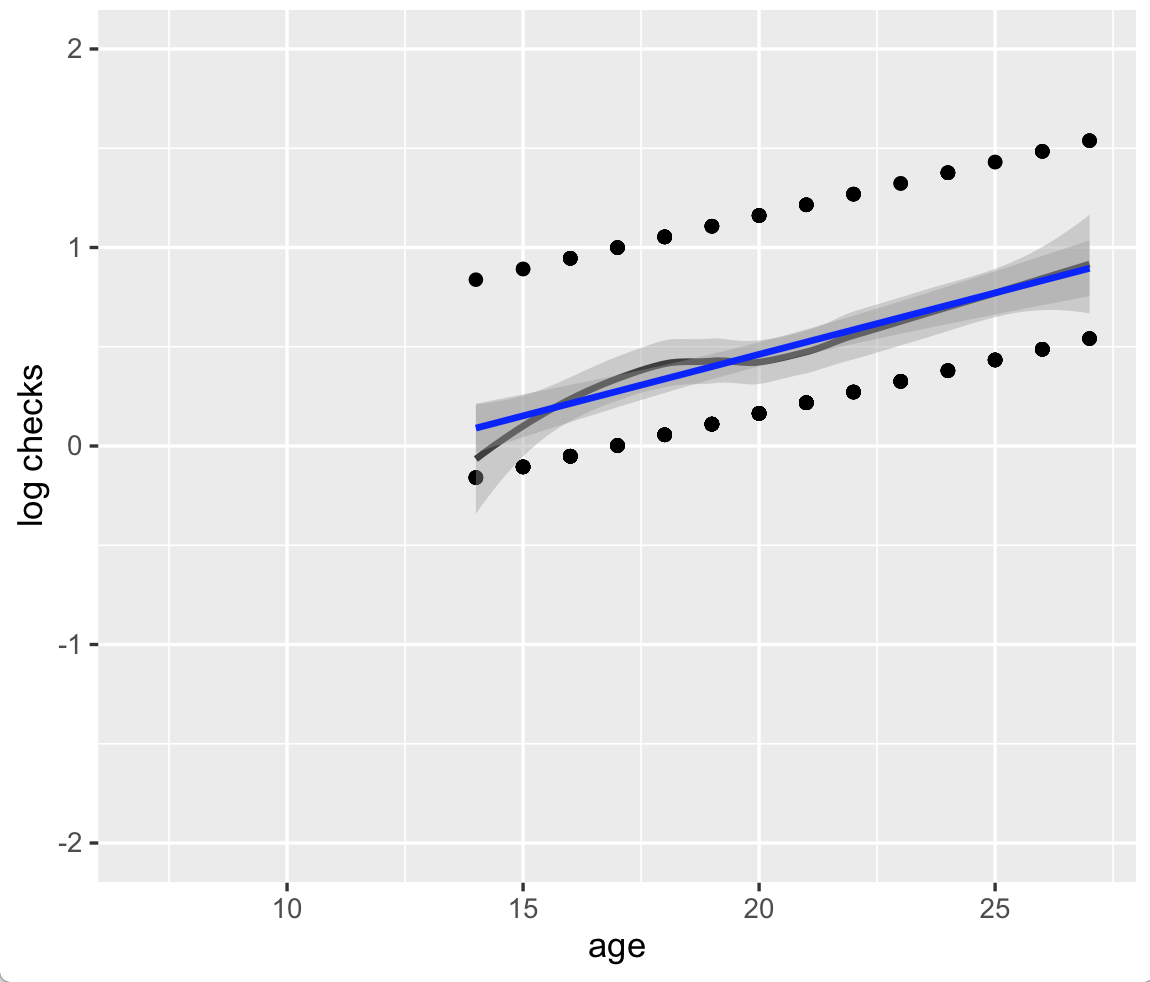
**Lab Assignment #10**

Due 1 week after your tenth lab session

Answer the following questions using the Arrests dataset and the Arrests Codebook.

For each question, provide your code and the answer (include all relevant plots).

1. Create a variable called “checksbinary” that equals 1 if an arrestee’s name appears in a police database for a previous arrest, conviction, or parole and 0 if their name does not appear
2. Create a subset of the Arrests data frame called Arrests2 that includes the following variables:
   1. checksbinary
   2. race
   3. age
3. Does the age variable adhere to the assumption of linearity?



From the above Q-Q plot, we can tell age adheres to the assumption of linearity with log checks.

1. Estimate a logistic regression model where the checksbinary variable is regressed on race and age.
   1. Interpret the coefficient for the race variable.

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.91340 0.45213 -2.020 0.04336 \*

age 0.05387 0.01914 2.815 0.00488 \*\*

race 0.99701 0.30664 3.251 0.00115 \*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

race: The logit coefficient is 0.997, which indicates that being black is associated with a .997 increase in log checks than being white.

**Appendix Code**

setwd("/Users/Leah/Downloads")

good <- read.csv("Arrests.csv",,header=TRUE, sep=",")

goodR<- na.omit(good[,c("released","race","age","gender","employed","citizen","checks")])

#Q1: Create a variable called “checksbinary” that equals 1 if an arrestee’s name appears in a police database for a previous arrest, conviction, or parole and 0 if their name does not appear

checksbinary <- ifelse(goodR$checks>0,1,0)

#Q2: Create a subset of the Arrests data frame called Arrests2 that includes the following variables: a. checksbinary, b. race, c. age

Arrests2 <- na.omit(goodR[,c("race","age")])

Arrests2["checksbinary"] <- NA

Arrests2$checksbinary <- checksbinary

#Q3: Check if the age variable adhere to the assumption of linearity

# First, estimate the logistic regression model with all variables

linearity <- glm(formula = checksbinary ~. , family=binomial(link='logit'),data= Arrests2)

# Next, use the predict function to produce logodds for plotting

logodds <- predict(linearity)

plotlin <- with(goodR, data.frame(age = age, logit = logodds))

ggplot(plotlin, aes(x = age, y = logit))+geom\_point()+labs(x = "age", y = "log checks") +geom\_smooth(method = "loess", col = "#3e3e3e")+geom\_smooth(method = "lm", col = "blue")+scale\_x\_continuous(limits=c(7, 27))+scale\_y\_continuous(limits= c(-2,2))

# Estimate the logistic regression model & the null model without predictors

lm1<-glm(checksbinary~age + race, family=binomial(link='logit'), data=goodR)

summary(lm1)