HW4\_567

Matthew Stoebe

2024-11-12

#a Use lm() to fit the multiple regression model described above. Provide estimates for the regression parameters β0, β1, β2 by showing the coefficients table from the summary output.

## Loading required package: carData

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -6.0646629 4.27194117 -1.419650 1.630896e-01  
## income 0.5987328 0.11966735 5.003310 1.053184e-05  
## education 0.5458339 0.09825264 5.555412 1.727192e-06

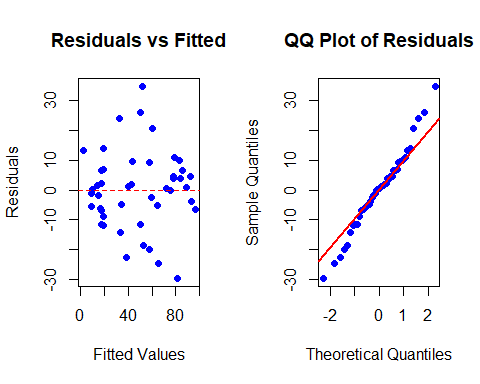
#b

confint(model, level = 0.95)

## 2.5 % 97.5 %  
## (Intercept) -14.6857892 2.5564634  
## income 0.3572343 0.8402313  
## education 0.3475521 0.7441158

#c

# Set up plotting area  
par(mfrow = c(1, 2)) # Two plots side by side  
  
# Residuals vs Fitted Values  
plot(model$fitted.values, resid(model),  
 main = "Residuals vs Fitted",  
 xlab = "Fitted Values",  
 ylab = "Residuals",  
 pch = 19, col = "blue")  
abline(h = 0, lty = 2, col = "red")  
  
# QQ Plot of Residuals  
qqnorm(resid(model), main = "QQ Plot of Residuals", pch = 19, col = "blue")  
qqline(resid(model), col = "red", lwd = 2)

 It appears that the variance of the residuals increases from left to right. We also see that for the outer points, the residuals fall off the line of normality. This implies that the assumptions are not satisfied

#d

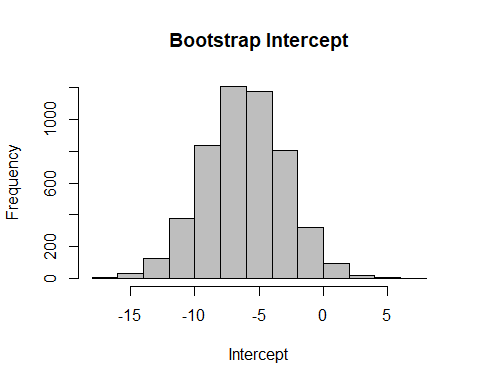
library(boot)

## Warning: package 'boot' was built under R version 4.4.2

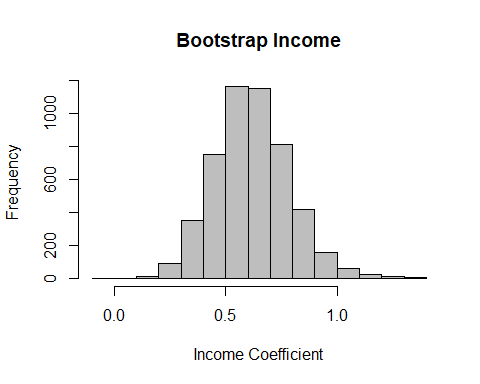
##   
## Attaching package: 'boot'

## The following object is masked from 'package:car':  
##   
## logit

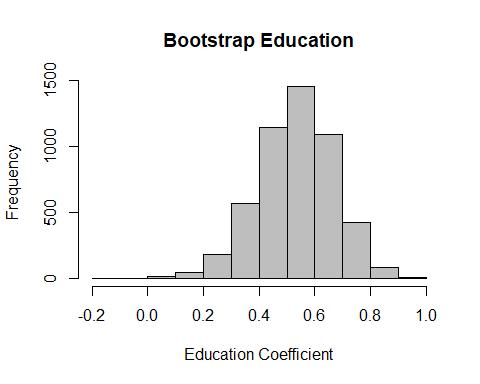
boot\_fn <- function(data, indices) {  
 d <- data[indices, ]   
 fit <- lm(prestige ~ income + education, data = d)  
 return(coef(fit))  
}  
  
set.seed(123)  
  
bootstrap\_results <- boot(data = Duncan, statistic = boot\_fn, R = 5000)  
  
boot\_estimates <- bootstrap\_results$t  
colnames(boot\_estimates) <- names(coef(model))  
  
# Histogram for Intercept  
hist(boot\_estimates[, "(Intercept)"], main = "Bootstrap Intercept",  
 xlab = "Intercept", col = "grey", border = "black")



# Histogram for Income  
hist(boot\_estimates[, "income"], main = "Bootstrap Income",  
 xlab = "Income Coefficient", col = "grey", border = "black")



# Histogram for Education  
hist(boot\_estimates[, "education"], main = "Bootstrap Education",  
 xlab = "Education Coefficient", col = "grey", border = "black")



#e

original\_estimates <- coef(model)  
bootstrap\_means <- colMeans(boot\_estimates)  
bias <- bootstrap\_means - original\_estimates  
  
print(bias)

## (Intercept) income education   
## -0.09718098 0.01898643 -0.01499215

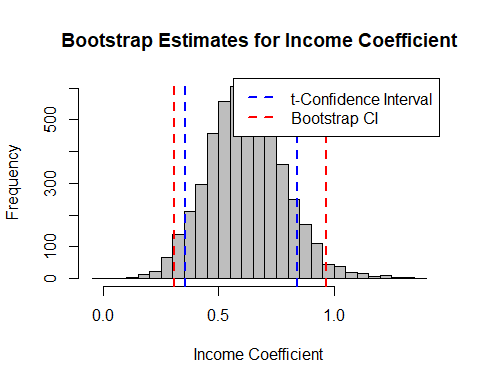
#f

percentile\_CI <- apply(boot\_estimates, 2, function(x) quantile(x, probs = c(0.025, 0.975)))  
  
# Transpose for better readability  
percentile\_CI <- t(percentile\_CI)  
colnames(percentile\_CI) <- c("2.5 %", "97.5 %")  
percentile\_CI

## 2.5 % 97.5 %  
## (Intercept) -12.3272482 -0.08135699  
## income 0.3075327 0.96596068  
## education 0.2474596 0.78284989

#g

# Reconstruct histogram for income bootstrap estimates  
hist(boot\_estimates[, "income"],   
 main = "Bootstrap Estimates for Income Coefficient",  
 xlab = "Income Coefficient",   
 col = "grey",   
 border = "black",   
 breaks = 30)  
  
# Add t-based confidence interval (from part b)  
t\_CI <- confint(model, level = 0.95)["income", ]  
abline(v = t\_CI, col = "blue", lwd = 2, lty = 2)  
  
# Add bootstrap confidence interval (from part f)  
boot\_CI\_income <- percentile\_CI["income", ]  
abline(v = boot\_CI\_income, col = "red", lwd = 2, lty = 2)  
  
legend("topright", legend = c("t-Confidence Interval", "Bootstrap CI"),  
 col = c("blue", "red"), lty = 2, lwd = 2)



Our bootstrap confidence interval is wider than the t confidence interval. this is expected