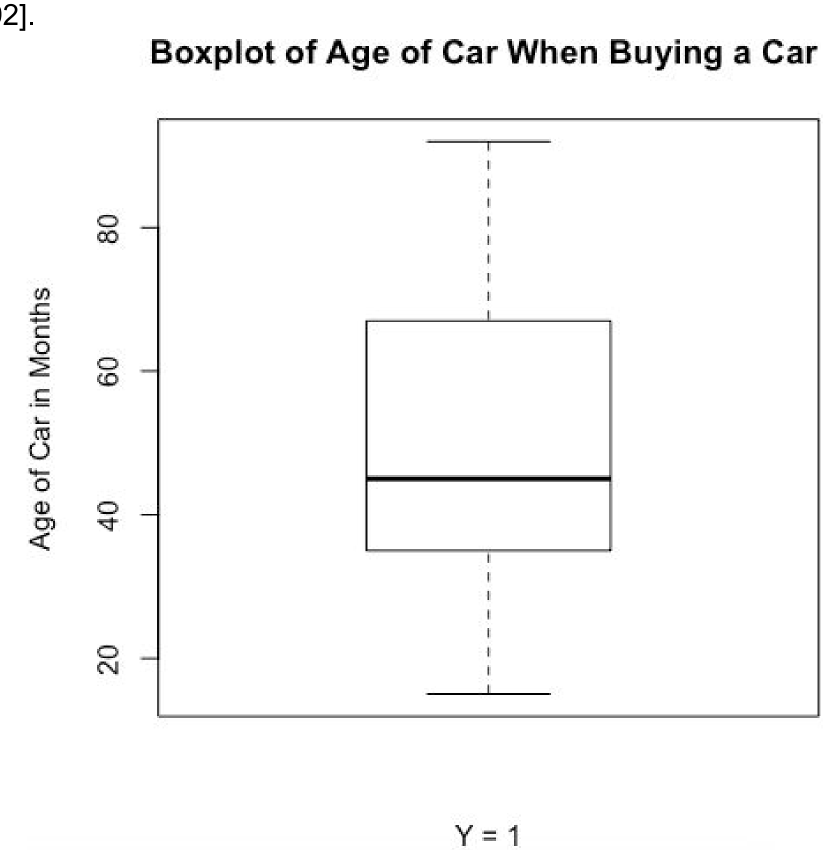
PROBLEM 1

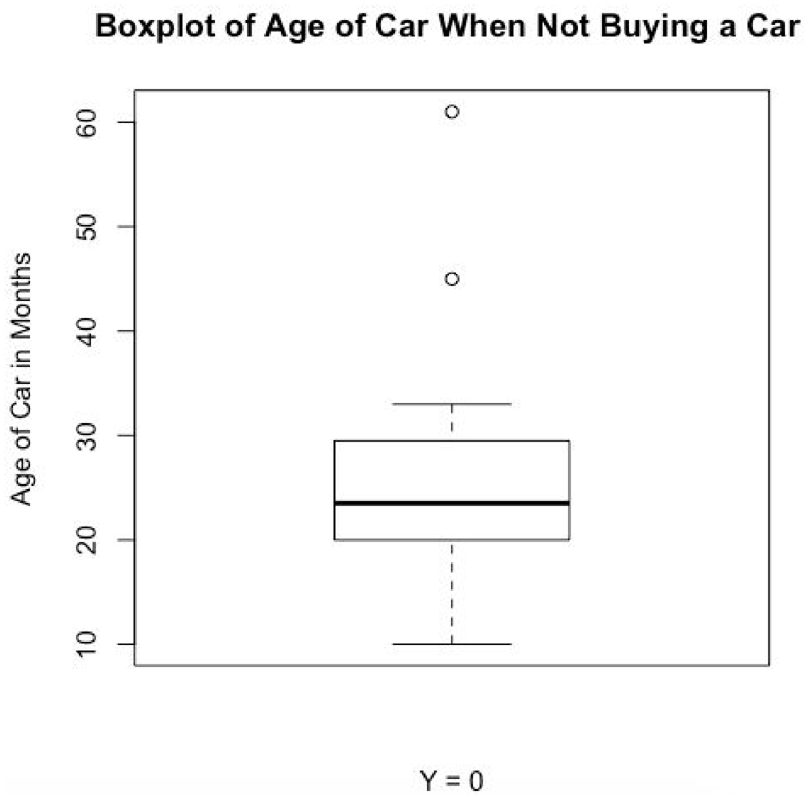
INTRODUCTION

Is there a relationship between buying a car and the age of the currently owned car? Data was gathered on customers who visited car dealerships to see if they bought a car and the age of the current car. All this information was recorded in “Car.csv”. The first column, Y, has the value 1 if the customer bought a new car and 0 if they did not buy a car. The second column X is the age of the current owned car in months. This data set is interesting and important because we can determine if there’s a relationship between buying a car and the age of the currently owned car. This will help dealerships predict the probability of a customer buying a car based on their current owned car’s age in month.

SUMMARY

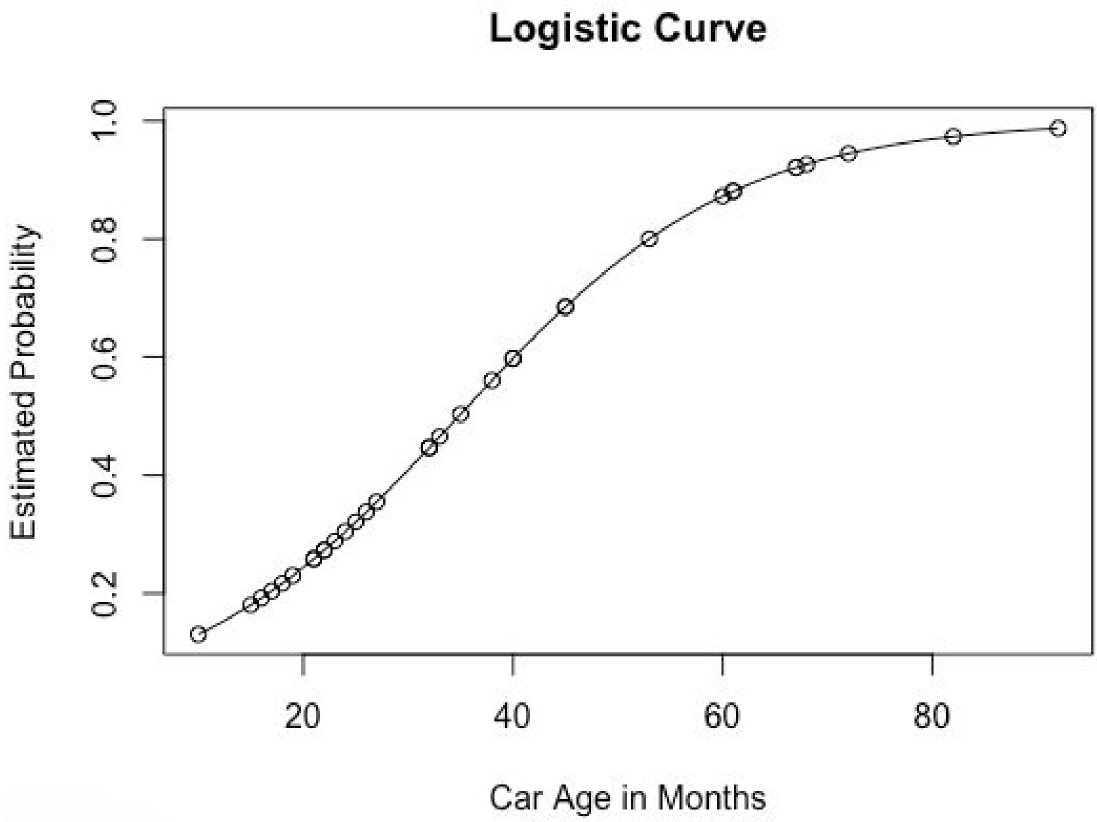
There are a total of 33 observations and 17 of those bought a car. The range of a car’s age in month is [10, 92].





This first box plot shows the range of the age of the currently owned car in months for customers who successfully purchased a car. It seems like the customer who bought a car had the youngest car age at 15 months and the oldest car age at 92 months. The median age of car for a customer to buy a new is car is around 45 months.

The second boxplot shows the range of age of cars in months for customers who did not buy a car when going into a car dealership. The youngest currently owned car is 10 months old whereas the oldest car age that is not the outliers is around 32 months old. However, this plot has 2 outliers at 45 months and 61 months old.



In the above logistic curve plot, there isn't a steep slope so it does not appear that the age of a car between [10, 20] and [60, 92] has a significant effect on whether a customer buys a car or not. However, the slope is steeper when the car age is between [20, 60] where there may be an effect on whether a customer buys a car or not.

ANALYSIS

Null Hypothesis: $B_1 = 0$
Alternative Hypothesis: $B_1 \neq 0$

X_1 = age of car in months
$\text{logit}(\pi(x)) = -2.65826 + 0.07635X_1$

The probability of 0.5 for a customer buying a car is when the age of the currently owned car is 34.815485 months old.

The value of $\exp(B_1)$ is 1.079343.

The Wald test-statistic for the null hypothesis is 2.64. Therefore, the Wald p-value is 0.008290531.

The Wald 99% confidence interval after exponential is (1.001858, 1.162822).

INTERPRETATION

We are testing to see if there's an effect on the age of car and a customer buying a new car.

exp(α) is an estimate of the odds of a customer buying a car when the age of the currently owned car is 0. This has no practical interpretation because when the age is 0, it exceeds our data's range. The minimum of our data's range for the age of car is 10 months. Also, a car of age 0 means the car is newly bought and that means the customer will not need to buy a new car again just yet.

The odds of buying a new car when the age of car increases by 1 month are $\exp(B_1) = 1.079343$ times what they were.

A customer has a 50% chance of a buying a new car if their currently owned car is 34.815485 months old.

Since alpha = 0.01 and our p-value is 0.008290531 which is less than alpha, we reject the null hypothesis and conclude that the age of car has some (positive) effect on a customer buying a new car. A p-value of 0.008290531 means that if in reality, there is no effect on the age of car on a customer buying a new car, the probability of observing our data or more extreme is 0.8290531%.

When the age of car increases by 1 month, the odds a customer buying a new car are between 1.001858 and 1.162822 times that of what they were with 99% confidence.

CONCLUSION

We tested to see if there was an effect on age of currently owned car in months on buying a new car when a customer goes to the car dealership. Since our 99% confidence interval for $\exp(B_1)$ did not contain 1, it suggests an influence on age of currently owned car on the odds of buying a car (Y = 1). Overall, the older the car, the higher the probability of buying a car. This makes sense since an older car means it is time for change and an upgrade by buying a new car.

APPENDIX PROBLEM 1

```
#import data
Car <- read.csv("~/Desktop/Car.csv")

zero = Car[Car$Y == 0, ]
one = Car[Car$Y == 1, ]

boxplot(one$X, xlab = "Y = 1", ylab = "Age of Car in Months", main = "Boxplot of Age of Car
When Buying a Car ")
boxplot(zero$X, xlab = "Y = 0", ylab = "Age of Car in Months", main = "Boxplot of Age of Car
When Not Buying a Car ")

#logistic model
logit.model =  glm(formula = Y ~ X, family = binomial(logit), data = Car)
summary(logit.model)

#basic statistics on data
tot = sum(Car$Y)
mini = min(Car$X)
maxi = max(Car$X)

#plot curve
plot(Car$X,logit.model$fitted.values, xlab = "Car Age in Months",ylab = "Estimated Probability",
main = "Logistic Curve")
curve(predict(logit.model, data.frame(X=x), type="response"), add=TRUE)

estimates =  summary(logit.model)$coefficients[,1] # A vector of only the estimates
SE =  summary(logit.model)$coefficients[,2] #A vector of only the Wald SE's

bestimate = exp(estimates[2])

#calculate B CI
alpha = 0.01
z.a.2 = qnorm(1-alpha/2)
upper.bounds = estimates +z.a.2*SE
lower.bounds = estimates -z.a.2*SE
Wald.CI = cbind(lower.bounds,upper.bounds)
Wald.CI

#exp(B) CI
lower = exp(Wald.CI[2,1])
upper = exp(Wald.CI[2,2])
```

```
(-coef(logit.model)[1]) / coef(logit.model)[2]

predict(logit.model,newdata = data.frame(X = 34.815485),type = "response")
```