

STA 4320 HW 1

Mary Robinson

September 18 2024

Question 1

Part a

Loading Data and Finding the First Six Rows

```
# loading data
library(ISLR2)
require(ISLR2)
```

You can insert more R code chunks

Part b

Regression Output

```
y = Auto$mpg
x = Auto$horsepower
reg = lm(y ~ x)
summary(reg)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.5710  -3.2592  -0.3435   2.7630  16.9240
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 39.935861   0.717499   55.66  <2e-16 ***
## x           -0.157845   0.006446  -24.49  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.906 on 390 degrees of freedom
## Multiple R-squared:  0.6059, Adjusted R-squared:  0.6049
## F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16
```

```
#could also do: attach(Auto) then summary(lm(mpg ~ horsepower))
```

Answers with words can be typed directly without a code chunk.

Part c

```
plot(x, y,  
      mail = "MPG and Horsepower",  
      pch = 16,  
      xlab = "Horsepower",  
      ylab = "MPG"  
)
```

```
## Warning in plot.window(...): "mail" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "mail" is not a graphical parameter
```

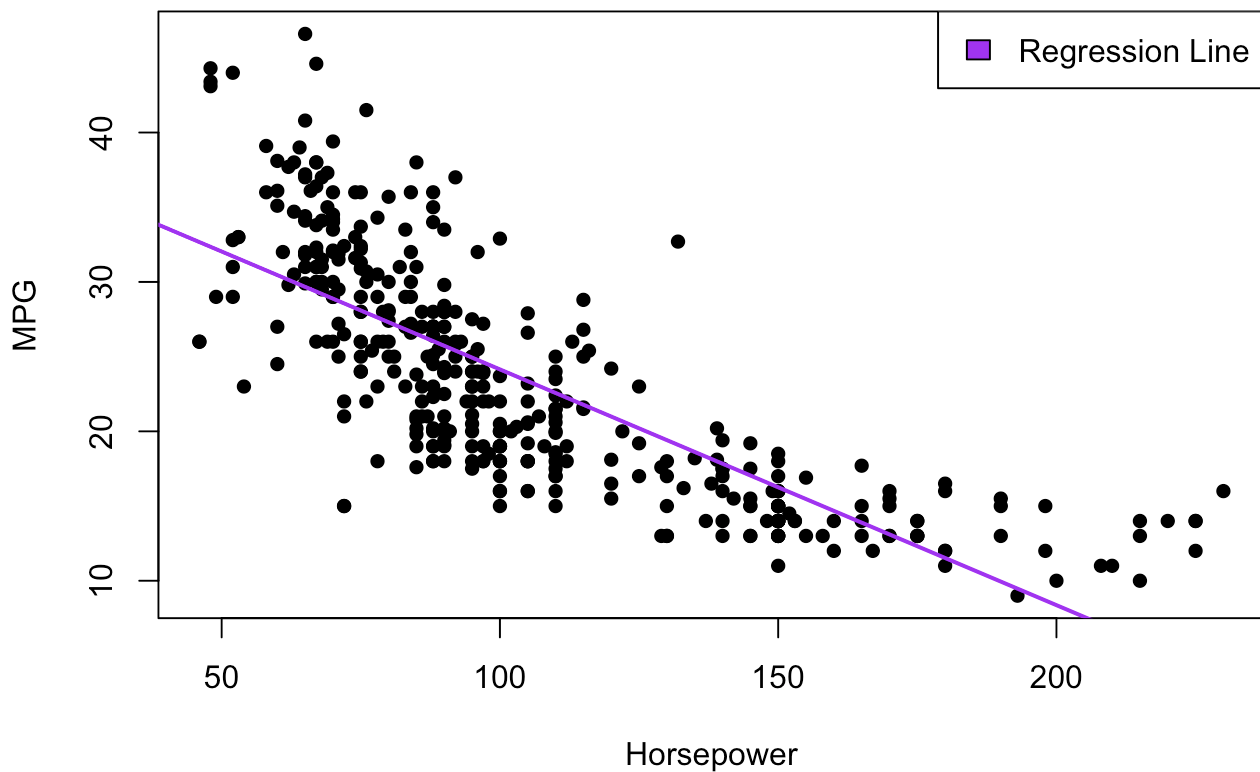
```
## Warning in axis(side = side, at = at, labels = labels, ...): "mail" is not a  
## graphical parameter  
## Warning in axis(side = side, at = at, labels = labels, ...): "mail" is not a  
## graphical parameter
```

```
## Warning in box(...): "mail" is not a graphical parameter
```

```
## Warning in title(...): "mail" is not a graphical parameter
```

```
#can also use Auto$Mpg or horsepower
```

```
abline(reg,  
       col = "purple",  
       lwd = 2)  
legend("topright",  
       legend = "Regression Line",  
       fill = "purple")
```



Part d

Yes. With the intercept already in the model:

$H_0: \beta_1 \geq 0$ vs $H_1: \beta_1 < 0$

The pvalue is less than e^{-16} . So we reject H_0 at the 0.05 level. #pvalue comes from our data t value and we see if it is less than our t value (see pic for details)

Part e

Yes this is the f test. We can look at the given F-statistic to compare our pvalue with it.

$H_0: \beta_0 = \beta_1 = 0$ vs (at least one of the β_0 or β_1 is nonzero)

The p value is less than $2.2e^{-16}$ so we reject H_0 at the 0.05 level.

part f

```
newx = data.frame(x = 98)

#simple point
round(as.numeric(predict(reg, newx)), 4)
```

```
## [1] 24.4671
```

```
#confidence interval  
round(predict(reg, newx, interval = "confidence", level = 0.95)[-1], 4)
```

```
## [1] 23.9731 24.9611
```

```
#prediction interval  
round(predict(reg, newx, interval = "prediction", level = 0.95)[-1], 4)
```

```
## [1] 14.8094 34.1248
```