Log-Linear Regression

July 26, 2023

MODEL 1: Base Model

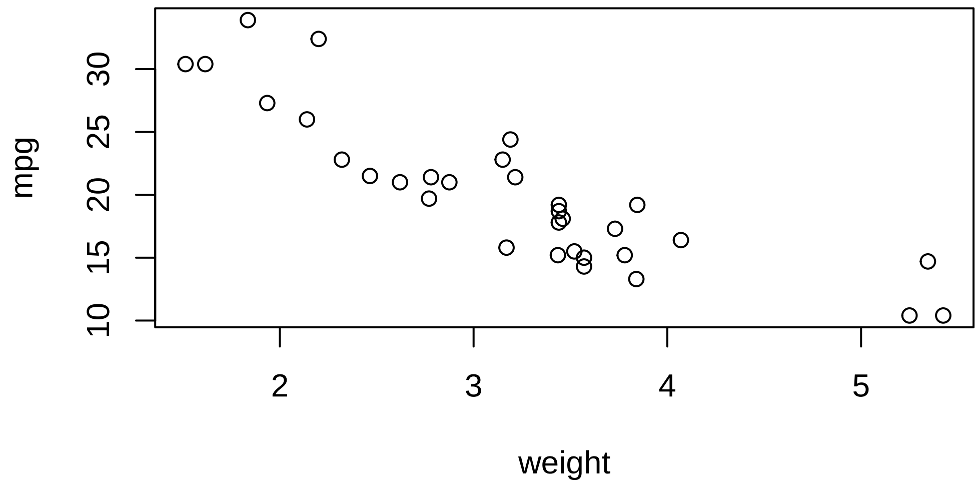
1. Reading the data

```
data(mtcars)
attach(mtcars)
```

2. Predict a mpg of car from their weight

Plot mpg and wt

```
plot(wt
     , mpg
     , xlab="weight", ylab="mpg")
```



3. Fitting a Linear Model

Simple Linear Regression

```
fit <- lm(mpg ~ wt, data = mtcars)
summary(fit)
```

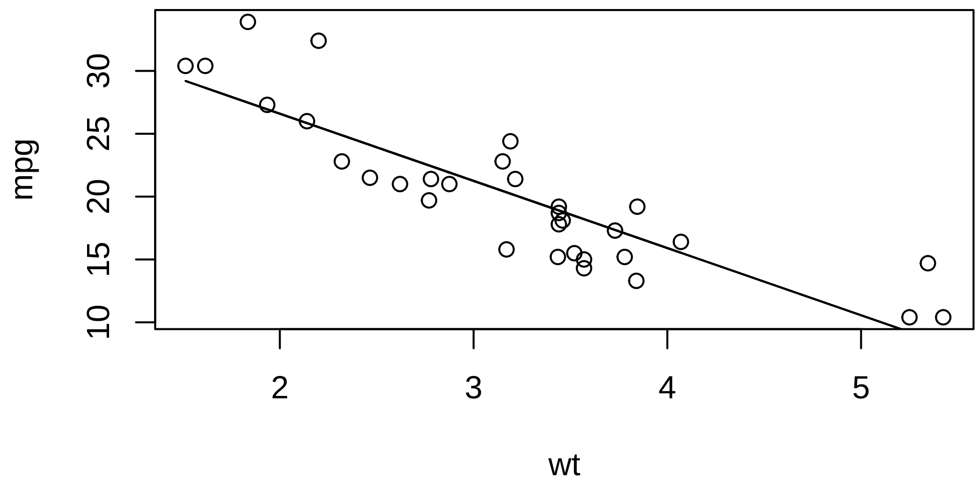
```
Call:
lm(formula = mpg ~ wt, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5432 -2.3647 -0.1252  1.4096  6.8727

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
wt           -5.3445     0.5591  -9.559 1.29e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.046 on 30 degrees of freedom
Multiple R-squared:  0.7528,    Adjusted R-squared:  0.7446
F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

```
plot(mtcars$wt,
     mtcars$mpg,
     main="",
     xlab="wt",
     ylab="mpg")
lines(mtcars$wt, fitted(fit))
```



Beta coefficients

```
fit$coefficients
```

```
(Intercept)          wt
  37.285126   -5.344472
```

Lists the predicted values in a fitted model

```
fitted(fit)
```

```
          Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive
          23.282611           21.919770           24.885952           20.102650
  Hornet Sportabout             Valiant          Duster 360           Merc 240D
          18.900144           18.793255           18.205363           20.236262
           Merc 230            Merc 280           Merc 280C          Merc 450SE
```

```
          20.450041           18.900144           18.900144           15.533127
         Merc 450SL         Merc 450SLC  Cadillac Fleetwood Lincoln Continental
          17.350247           17.083024            9.226650            8.296712
  Chrysler Imperial            Fiat 128         Honda Civic      Toyota Corolla
           8.718926           25.527289           28.653805           27.478021
      Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
          24.111004           18.472586           18.926866           16.762355
   Pontiac Firebird           Fiat X1-9       Porsche 914-2        Lotus Europa
          16.735633           26.943574           25.847957           29.198941
     Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
          20.343151           22.480940           18.205363           22.427495
```

Lists the residual values in a fitted model

Residuals are the vertical distances between the data and the fitted line. The Ordinary Least Squares (OLS) method minimizes the residuals. In OLS, the accuracy of a line through the sample data points is measured by the sum of squared residuals, and the goal is to make this sum as small as possible.

```
residuals(fit)
```

```
          Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive
         -2.2826106          -0.9197704          -2.0859521           1.2973499
  Hornet Sportabout             Valiant          Duster 360           Merc 240D
         -0.2001440          -0.6932545          -3.9053627           4.1637381
           Merc 230            Merc 280           Merc 280C          Merc 450SE
          2.3499593           0.2998560          -1.1001440           0.8668731
         Merc 450SL         Merc 450SLC  Cadillac Fleetwood Lincoln Continental
         -0.0502472          -1.8830236           1.1733496           2.1032876
  Chrysler Imperial            Fiat 128         Honda Civic      Toyota Corolla
          5.9810744           6.8727113           1.7461954           6.4219792
      Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
         -2.6110037          -2.9725862          -3.7268663          -3.4623553
   Pontiac Firebird           Fiat X1-9       Porsche 914-2        Lotus Europa
          2.4643670           0.3564263           0.1520430           1.2010593
     Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
         -4.5431513          -2.7809399          -3.2053627          -1.0274952
```

Statistical Significance and p-values

- The regression coefficient (3.45) is significantly dfferent from zero ($p < 0.001$)

- There is an expected increase of 3.45 lbs of weight for every 1 inch increases in height.

Multiple R-squred

```
# extacting the coefficient of determination
summary(fit)$r.squared
```

```
[1] 0.7528328
```

Compute Confidence Interval for Linear Regression

```
newdata = data.frame(wt = 4)
predict(fit, newdata, interval = "confidence")
```

```
       fit      lwr      upr
1 15.90724 14.49018 17.32429
```

F-Statistic

The F-statistic tests whether the predictor variables, taken together,predict the response variable.

MODEL 2: Log-Linear Model

```
fit2 <- lm(log(mpg) ~ wt, data = mtcars)
summary(fit2)
```

```
Call:
lm(formula = log(mpg) ~ wt, data = mtcars)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.210346 -0.085932 -0.006136  0.061335  0.308623 
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.83191    0.08396   45.64  < 2e-16 ***
wt          -0.27178    0.02500  -10.87 6.31e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1362 on 30 degrees of freedom
Multiple R-squared:  0.7976,	Adjusted R-squared:  0.7908 
F-statistic: 118.2 on 1 and 30 DF,  p-value: 6.31e-12
```

The beta estimate for wt in the above regression is -0.27178.

This means that for a one-unit increase in the weight of a car, the expected value of the log of its miles per gallon (mpg) is expected to decrease by -0.27178 units, after controlling for other variables in the model.

Since the response variable (log(mpg)) is on a logarithmic scale, the beta estimate for wt indicates a multiplicative effect on the mpg, rather than an additive effect.

```r
exp(-0.27178)
```

```
[1] 0.7620219
```

Specifically, we can interpret the beta estimate as suggesting that the expected ratio of mpg for two cars that differ by one unit in weight is e^(-0.27178) = 0.7620219

In other words, we would expect the car with the higher weight to have a fuel efficiency that is approximately 76.2% lower than the car with the lower weight, after controlling for other factors in the model.

MODEL 3a: Another Linear-Linear Model

```r
# Convert am to a factor variable
mtcars$am <- factor(mtcars$am, labels = c("Automatic", "Manual"))

fit3a <- lm((mpg) ~ wt + am, data = mtcars)
summary(fit3a)
```

```
Call:
lm(formula = (mpg) ~ wt + am, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5295 -2.3619 -0.1317  1.4025  6.8782 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 37.32155    3.05464  12.218 5.84e-13 ***
wt          -5.35281    0.78824  -6.791 1.87e-07 ***
amManual    -0.02362    1.54565  -0.015    0.988    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.098 on 29 degrees of freedom
Multiple R-squared:  0.7528,    Adjusted R-squared:  0.7358 
F-statistic: 44.17 on 2 and 29 DF,  p-value: 1.579e-09
```

MODEL 3b: Another Log-Linear Model

```
fit3b <- lm(log(mpg) ~ wt + am, data = mtcars)
summary(fit3b)
```

```
Call:
lm(formula = log(mpg) ~ wt + am, data = mtcars)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.21351 -0.08281  0.00304  0.04962  0.32349 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.89834    0.13567  28.734  < 2e-16 ***
wt          -0.28699    0.03501  -8.198 4.87e-09 ***
amManual    -0.04307    0.06865  -0.627    0.535    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1376 on 29 degrees of freedom
```

```
Multiple R-squared:  0.8003,    Adjusted R-squared:  0.7865
F-statistic:  58.1 on 2 and 29 DF,  p-value: 7.186e-11
```

```r
exp(-0.04307)
```

```
[1] 0.9578443
```

```r
exp(-0.2869)
```

```
[1] 0.7505868
```

The beta estimate of amManual in the above regression is -0.04307. This means that, after controlling for weight, the expected value of the log of the miles per gallon (mpg) for a car with a manual transmission is expected to be -0.04307 units lower than for a car with an automatic transmission.

Since the am variable was converted to a factor variable, the beta estimate for amManual represents the difference in the expected value of the response variable between cars with manual and automatic transmissions, after controlling for weight. In this case, the beta estimate indicates that cars with manual transmissions are expected to have a higher fuel efficiency than cars with automatic transmissions, all else being equal.

It's important to note that the p-value for the amManual coefficient is greater than the conventional threshold of 0.05 for statistical significance. This means that we cannot reject the null hypothesis that the amManual coefficient is equal to zero, and thus the difference in fuel efficiency between manual and automatic transmissions may not be statistically significant. However, this should be interpreted with caution, as it's possible that the sample size or other factors in the data are affecting the p-value.