1 Review

```
model_mpg <- lm(mpg ~ hp + am + wt, data = mtcars)

summary(model_mpg)
```

```
##
## Call:
## lm(formula = mpg ~ hp + am + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## am           2.083710   1.376420   1.514 0.141268
## wt          -2.878575   0.904971  -3.181 0.003574 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

```
anova(model_mpg) #anova stands for analysis of variance
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## hp         1 678.37  678.37 105.354 5.395e-11 ***
## am         1 202.24  202.24  31.408 5.335e-06 ***
## wt         1  65.15   65.15  10.118  0.003574 **
## Residuals 28 180.29    6.44
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
678.37 + 202.24 + 65.15 + 180.29
```

```
## [1] 1126.05
```

```r
mtcars %>%
  mutate(error = mpg - mean(mpg)) %>%
  mutate(error_sq = error^2) %>%
  glimpse()
```

```
## Observations: 32
## Variables: 13
## $ mpg      <dbl> 21.0, 21.0, 22.8, 21.4, 18.7, 18.1, 14.3, 24.4, 22.8,...
## $ cyl      <dbl> 6, 6, 4, 6, 8, 6, 8, 4, 4, 6, 6, 8, 8, 8, 8, 8, 8, 4,...
## $ disp     <dbl> 160.0, 160.0, 108.0, 258.0, 360.0, 225.0, 360.0, 146....
## $ hp       <dbl> 110, 110, 93, 110, 175, 105, 245, 62, 95, 123, 123, 1...
## $ drat     <dbl> 3.90, 3.90, 3.85, 3.08, 3.15, 2.76, 3.21, 3.69, 3.92,...
## $ wt       <dbl> 2.620, 2.875, 2.320, 3.215, 3.440, 3.460, 3.570, 3.19...
## $ qsec     <dbl> 16.46, 17.02, 18.61, 19.44, 17.02, 20.22, 15.84, 20.0...
## $ vs       <dbl> 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1,...
## $ am       <dbl> 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,...
## $ gear     <dbl> 4, 4, 4, 3, 3, 3, 3, 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 4,...
## $ carb     <dbl> 4, 4, 1, 1, 2, 1, 4, 2, 2, 4, 4, 3, 3, 3, 4, 4, 4, 1,...
## $ error    <dbl> 0.909375, 0.909375, 2.709375, 1.309375, -1.390625, -1...
## $ error_sq <dbl> 0.8269629, 0.8269629, 7.3407129, 1.7144629, 1.9338379...
```

```r
mtcars %>%
  mutate(error = mpg - mean(mpg)) %>%
  mutate(error_sq = error^2) %>%
  summarize(sum(error_sq))
```

```
##   sum(error_sq)
## 1      1126.047
```

```r
summary(model_mpg)$r.squared
```

```
## [1] 0.8398903
```

```r
(678.37 + 202.24 + 65.15)/(678.37 + 202.24 + 65.15 + 180.29)
```

```
## [1] 0.8398917
```

2 Binomial Likelihood Example

You have a friend who claims that they can identify caffeinated and non-caffeniated coffee. Let their success rate be an unknown parameter p. You do a double blind test with 10 cups of coffee. Your friend identifies 8 cups of coffee correctly. Which of the following is the most likely estimation for p?

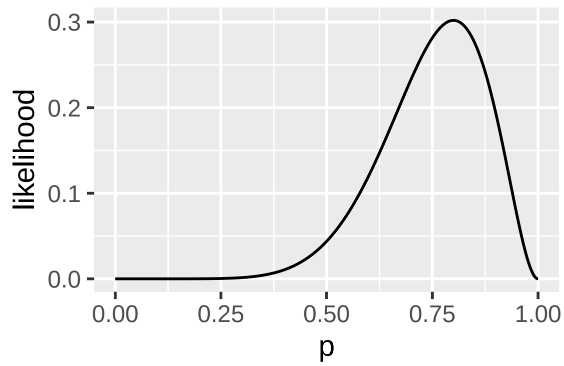
a) 0.3

b) 0.5

c) 0.8

The answer may seem intuitive but let's see the reasoning behind it

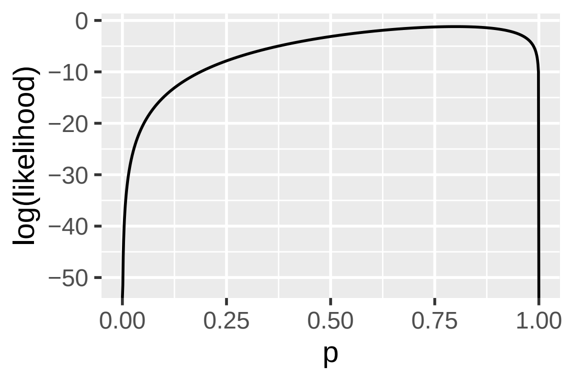
```
p <- seq(from = 0, to = 1, by = 0.001)
likelihood <- dbinom(x = 8, size = 10, prob = p)

data<- as.data.frame(cbind(p, likelihood))

data %>%
  ggplot(aes(x = p, y = likelihood)) +
  geom_line()
```




```
data %>%
filter(likelihood == max(likelihood))
```

```
##     p likelihood
## 1 0.8  0.3019899
```

```
data %>%
  ggplot(aes(x = p, y = log(likelihood))) +
  geom_line()
```




```
data %>%
filter(log(likelihood) == max(log(likelihood)))
```

```
##     p likelihood
## 1 0.8  0.3019899
```

Estimating p is quite easy when we are only testing one friend. We are now going to consider two scenarios.

The experiment is repeated with a friend who tests 50 cups of coffee and correctly identifies 40 cups. What is the estimate for p?

In second scenario:

You repeat the experiment with five friends where each friend tests 10 cups of coffee. Two friends are correct about 7 cups, 1 friend is correct about 8 cups, 2 friends are correct about 9 cups. What is the estimate for p?

3 Binomial Likelihood Using Calculus