

Data 609 Assignment #1

Joseph Simone

2/9/2020

```
library(ggplot2)
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(knitr)
library(grid)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##   combine
```

```
library(latex2exp)
```

Chapter 1

Page 17, #9

The data in the accompanying table show the speed n (in increments of 5 mph) of an automobile and the associated distance a_n in feet required to stop it once the brakes are applied. For instance, $a_6 = 6 \times 5 = 30$ requires a stopping distance of $a_6 = 47 \text{ ft}$.

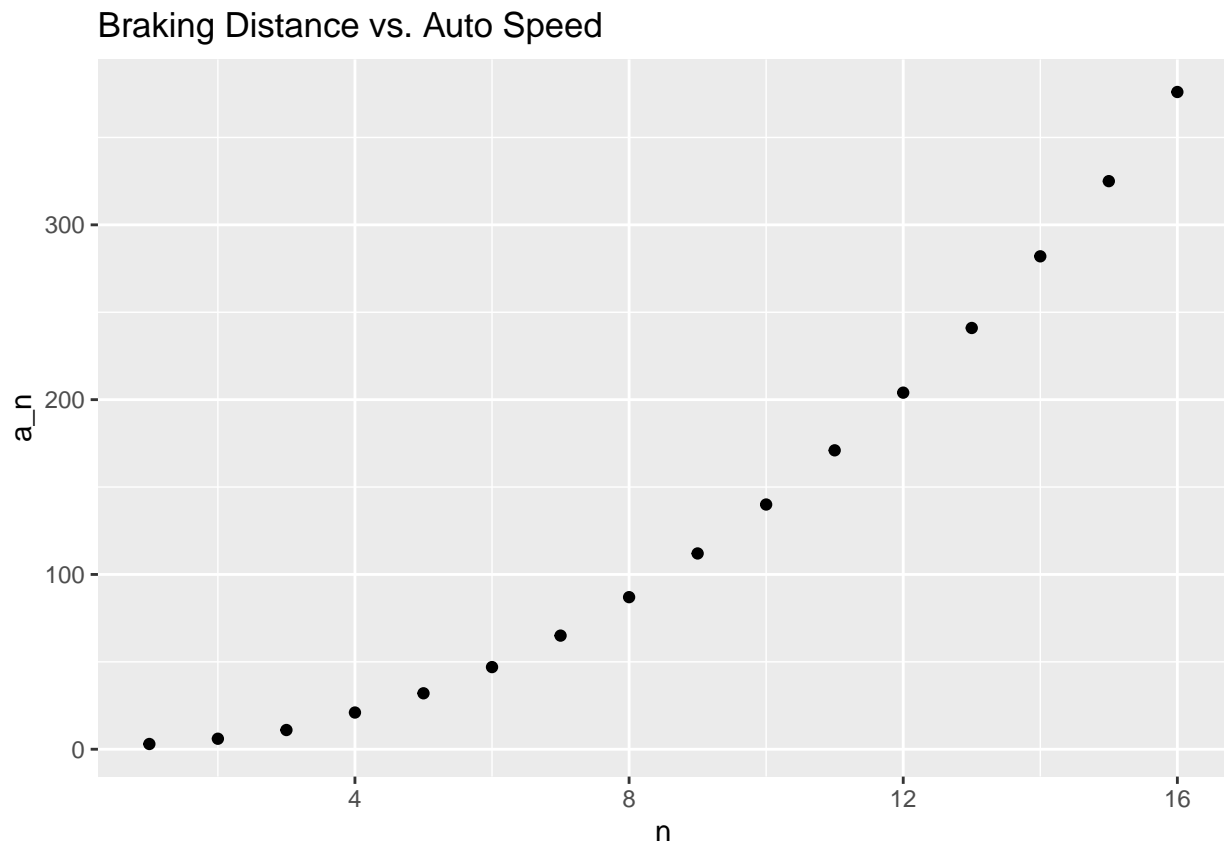
```
n <- (seq(1, 16))
a_n <- c(3, 6, 11, 21, 32, 47, 65, 87, 112, 140, 171, 204, 241, 282, 325, 376)
mydf <- data.frame(n, a_n)
mydf
```

```
##      n a_n
## 1    1   3
## 2    2   6
## 3    3  11
## 4    4  21
## 5    5  32
## 6    6  47
## 7    7  65
## 8    8  87
## 9    9 112
## 10   10 140
## 11   11 171
## 12   12 204
## 13   13 241
## 14   14 282
## 15   15 325
## 16   16 376
```

Part A. Calculate and plot the change Δa_n versus n . Does the graph reasonably approximate a linear relationship?

The relationship between braking distance and speed:

```
ggplot(mydf, aes(x = n, y = a_n)) + geom_point() + ggtitle("Braking Distance vs. Auto Speed")
```



The relationship between these variables appears to be quadratic.

The relationship between the change in braking distance and speed:

```
mydf <- mydf %>% mutate(delta_a = lead(a_n) - a_n)
mydf
```

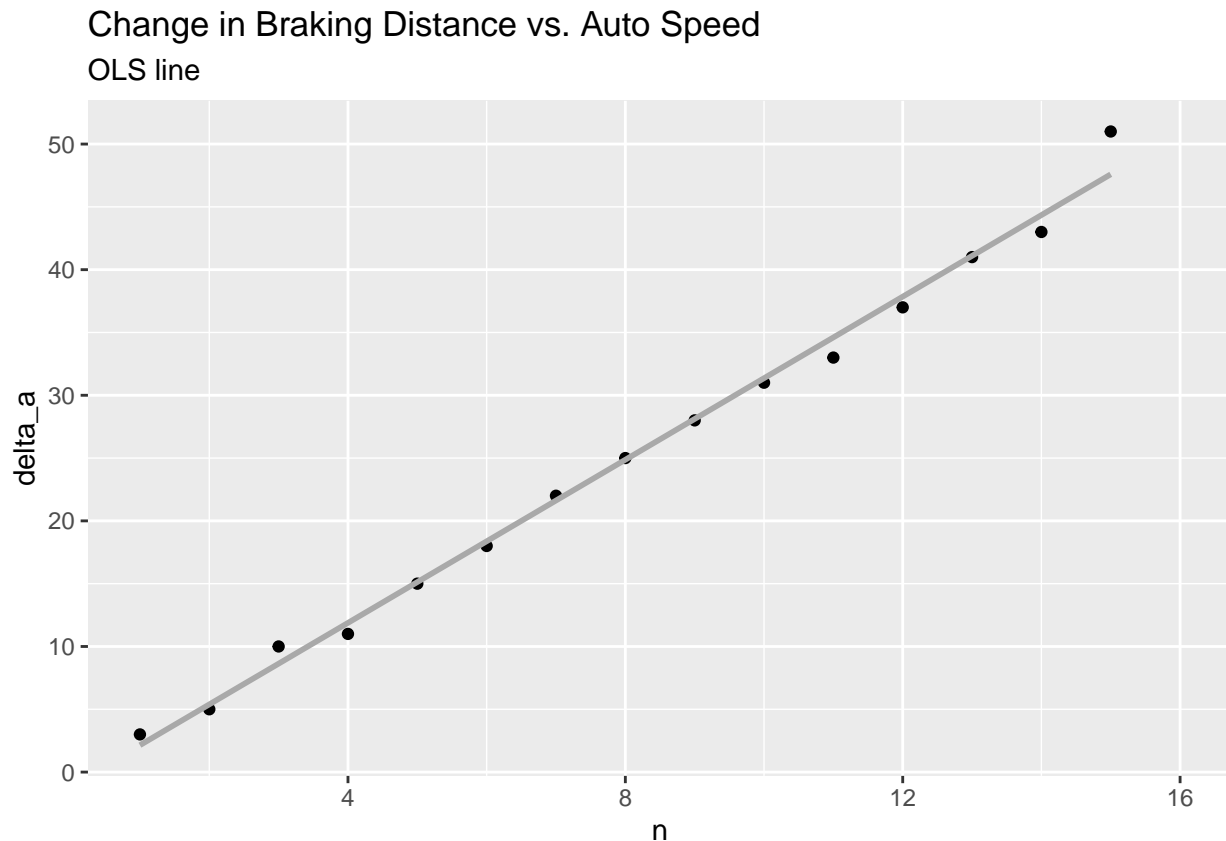
```
##      n a_n delta_a
## 1     1   3        3
## 2     2   6        5
## 3     3  11       10
## 4     4  21       11
## 5     5  32       15
## 6     6  47       18
## 7     7  65       22
## 8     8  87       25
## 9     9 112       28
## 10    10 140       31
## 11    11 171       33
## 12    12 204       37
## 13    13 241       41
## 14    14 282       43
## 15    15 325       51
## 16    16 376      NA
```

```
g <- ggplot(mydf, aes(x = n, y = delta_a))
g <- g + geom_point() + ggtitle("Change in Braking Distance vs. Auto Speed")
```

```
g <- g + labs(subtitle = "OLS line")
g + geom_smooth(method = "lm", color = "darkgray", se = FALSE)
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



Part B. Based on your conclusions in part (a), find a difference equation model for the stopping distance data. Test your model by plotting the errors in the predicted values against n . Discuss the appropriateness of the model.

The Coefficient and Intercept from the OLS model:

```
mylm <- lm(delta_a ~ n, data = mydf)
mylm$coefficients
```

```
## (Intercept)          n
##   -1.104762    3.246429
```

Therefore, the Difference Equation:

$$\Delta a_n = 3.246n - 1.105$$

or

$$a_{n+1} = a_n + 3.246n - 1.105$$

Using the difference equation model to predict braking distances for speed increments 2 through 16.

```

pred_brake_dist <- function(n, a_n, slope, intercept) {
  a_n + slope * n + intercept
}

n <- 1
a_n <- 3

pred_vector <- 3

for (i in seq(2, 16)) {
  pred_vector[i] <- pred_brake_dist(n, a_n, mylm$coefficients[2], mylm$coefficients[1])
  n <- n + 1
  a_n <- pred_vector[i]
}

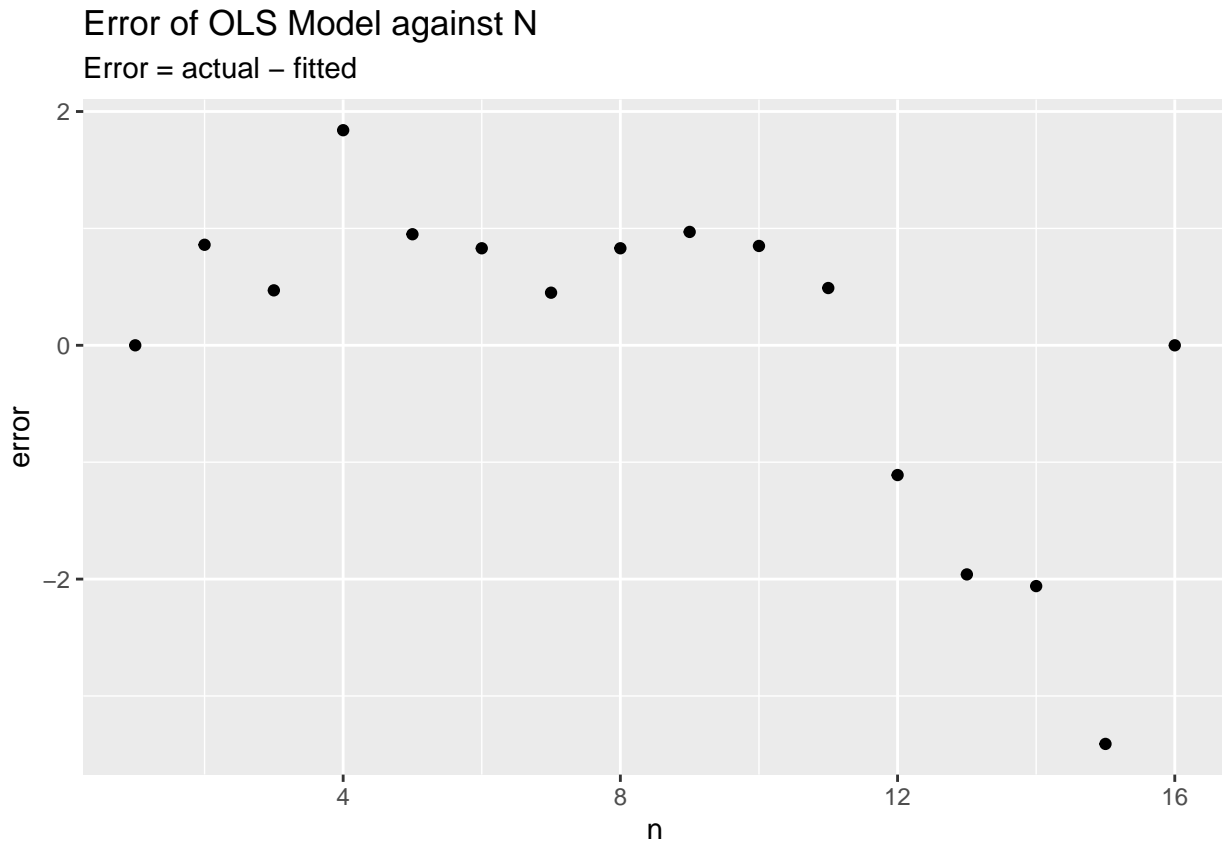
mydf["pred_a_n"] <- round(pred_vector, 2)

mydf$error <- mydf$a_n - mydf$pred_a_n

e <- ggplot(mydf, aes(x = n, y = error)) + geom_point()

e <- e + ggtitle("Error of OLS Model against N")
e + labs(subtitle = "Error = actual - fitted")

```



In the above plot, the error terms do not appear to be entirely random. Consequently, they appear to exhibit serial correlation. Therefore, one would have to be careful when interpreting the OLS estimates of the standard errors. These estimates are likely understated., considering performing formal tests for serial correlation (e.g. Durbin Watson).

Page 55, #6

An economist is interested in the variation of the price of a single product. It is observed that a high price for the product in the market attracts more suppliers. However, increasing the quantity of the product supplied tends to drive the price down. Over time, there is an interaction between price and supply. The economist has proposed the following model, where P_n represents the price of the product at year n , and Q_n represents the quantity. Find the equilibrium values for this system.

$$P_{n+1} = P_n - 0.1(Q_n - 500)$$

$$Q_{n+1} = Q_n + 0.2(P_n - 100)$$

To find the equilibrium solution to this equation, we set $P = P_{n+1} = P_n$ and $Q = Q_{n+1} = Q_n$, then rearrange the 2×2 linear system as follows:

$$\begin{bmatrix} 0 & -0.1 \\ 0.2 & 0 \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} -50 \\ 20 \end{bmatrix}$$

```
A <- matrix(c(0, -0.1, 0.2, 0), 2, 2, byrow = TRUE)
A_inv <- solve(A)
b <- c(-50, 20)
A_inv %*% b
```

```
##      [,1]
```

```
## [1,] 100
## [2,] 500
```

Part A. Does the model make sense intuitively? What is the significance of the constants 100 and 500? Explain the significance of the signs of the constants -0.1 and 0.2.

The model has an intuitive interpretation.

Price and quantity interact as follows:

A high (low) quantity has a negative (positive) impact on price.

A high (low) price has a positive (negative) influence on quantity.

The constant 100 in the second equation means that when the price is above 100, the quantity supplied will increase in the subsequent period. Consequently, whenever the price is below \$100, the quantity will decrease in the subsequent period.

In addition, the constant 500 in the first equation means that when the quantity supplied is greater than 500, the price will increase in the next period. Whenever the quantity is less than 500, the price will decrease in the subsequent period.

The multiplicative constant -0.1 in the first equation refers to the sensitivity of price to changes in quantity. The model assumes that price decreases by \$0.10 for every additional unit supplied.

The 0.2 constant in the second equation is used to describe the sensitivity of quantity to changes in price. In this model, the quantity supplied increases by 0.2 units for every \$1 increase in price.

Part B Test the initial conditions in the following table and predict the long-term behavior.

```
case <- c("A", "B", "C", "D")
price <- c(100, 200, 100, 100)
quantity <- c(500, 500, 600, 400)
mydf2 <- data.frame(case = case, price = price, quantity = quantity)
mydf2
```

```
##   case price quantity
## 1    A   100     500
## 2    B   200     500
## 3    C   100     600
## 4    D   100     400
```

```
price_calc <- function(p, q) p - 0.1 * (q - 500)
```

```
quantity_calc <- function(p, q) q + 0.2 * (p - 100)
```

```
# loop through all cases
for (i in seq(1, 4)) {
  p <- price[i]
  q <- quantity[i]
  model_df <- data.frame(time = 0, price = p, quantity = q)
```

```
# loop through multiple time periods
```

```

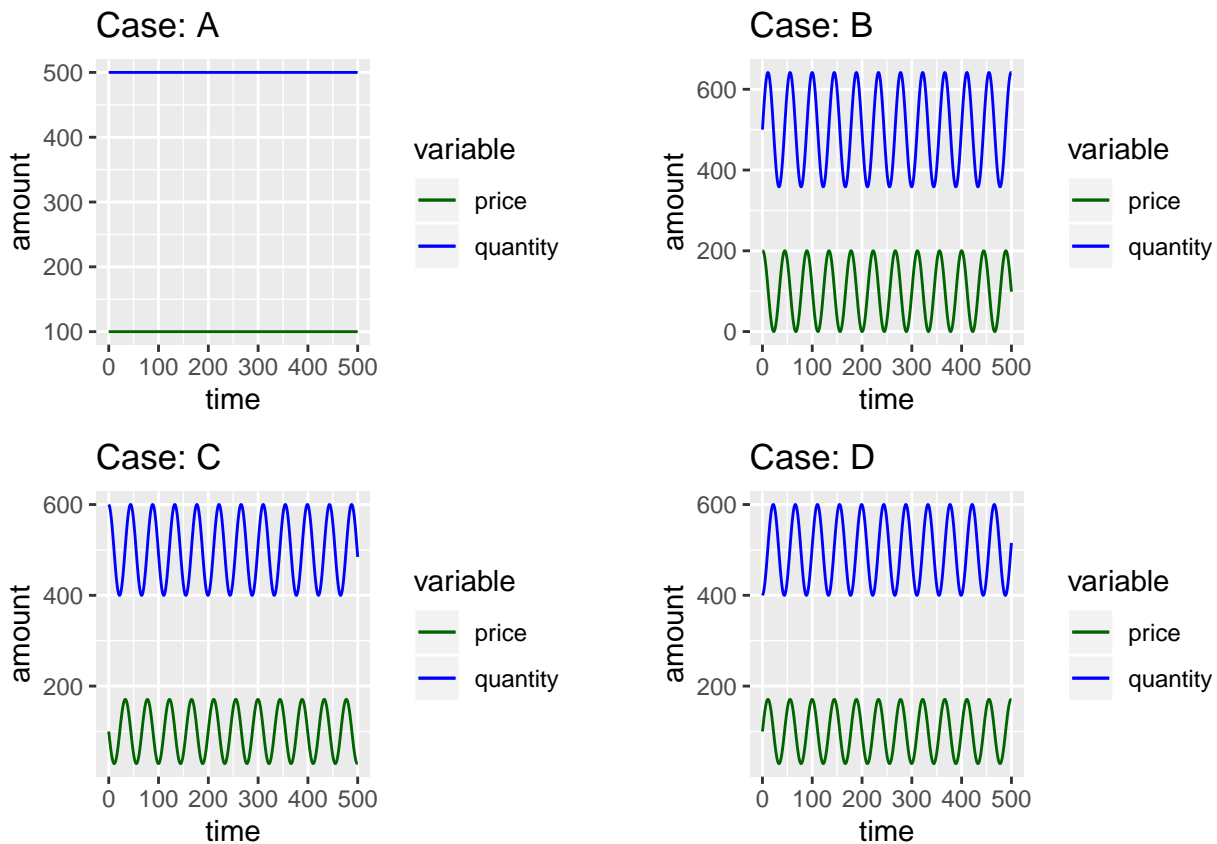
for (pd in seq(1, 500)) {
  p <- price_calc(p, q)
  q <- quantity_calc(p, q)
  model_df[pd + 1, ] <- c(pd, p, q)
}

g <- ggplot(data = model_df)
g <- g + geom_line(aes(x = time, y = quantity, color = "quantity"))
g <- g + geom_line(aes(x = time, y = price, color = "price"))
g <- g + ylab("amount")
g <- g + scale_colour_manual("variable", values = c(quantity = "blue", price = "darkgreen"))
g <- g + ggtitle(paste("Case:", case[i]))

assign(paste0("Case", i), g)
}

blank <- rectGrob(gp = gpar(col = "white"))
grid.arrange(Case1, blank, Case2, Case3, blank, Case4, ncol = 3, nrow = 2, widths = c(5,
0.5, 5))

```



-Case A: No change, as both price and quantity are at their equilibrium values of \$100 and 500, respectively.

-Case B: Price is above its equilibrium value; so quantity will increase. Price will then grow more slowly and eventually decrease in response to the quantity increase. Quantity will grow more slowly as price decreases,

and eventually decrease. When quantity falls below 500, price will begin to rebound. This results in a never-ending tug-of-war between price and quantity. In summary, both price and quantity exhibit oscillating behavior.

-Case C: Quantity is above its equilibrium value; so price will react negatively. Quantity grows more slowly and then falls as price dips below equilibrium value. But price eventually rebounds as quantity falls below 500. Like Case B, this scenario involves oscillatory behavior among both variables.

-Case D: Quantity is below equilibrium value; so price increases. The price increase spawns additional increases in quantity until quantity exceeds equilibrium. At this point, prices decrease, which results to a slowing increase, and eventually, a decrease in quantity. Once again, both variables exhibit oscillatory behavior.

Chapter 2

Page 79, #11

Determine whether the data set supports the stated proportionality model:

$$y \propto x^3$$

```
y <- c(0, 1, 2, 6, 14, 24, 37, 58, 82, 114)
x <- seq(1, 10)
```

```
mydf3 <- data.frame(y = y, x = x)
mydf3
```

```
##      y  x
## 1    0  1
## 2    1  2
## 3    2  3
## 4    6  4
## 5   14  5
## 6   24  6
## 7   37  7
## 8   58  8
## 9   82  9
## 10 114 10
```

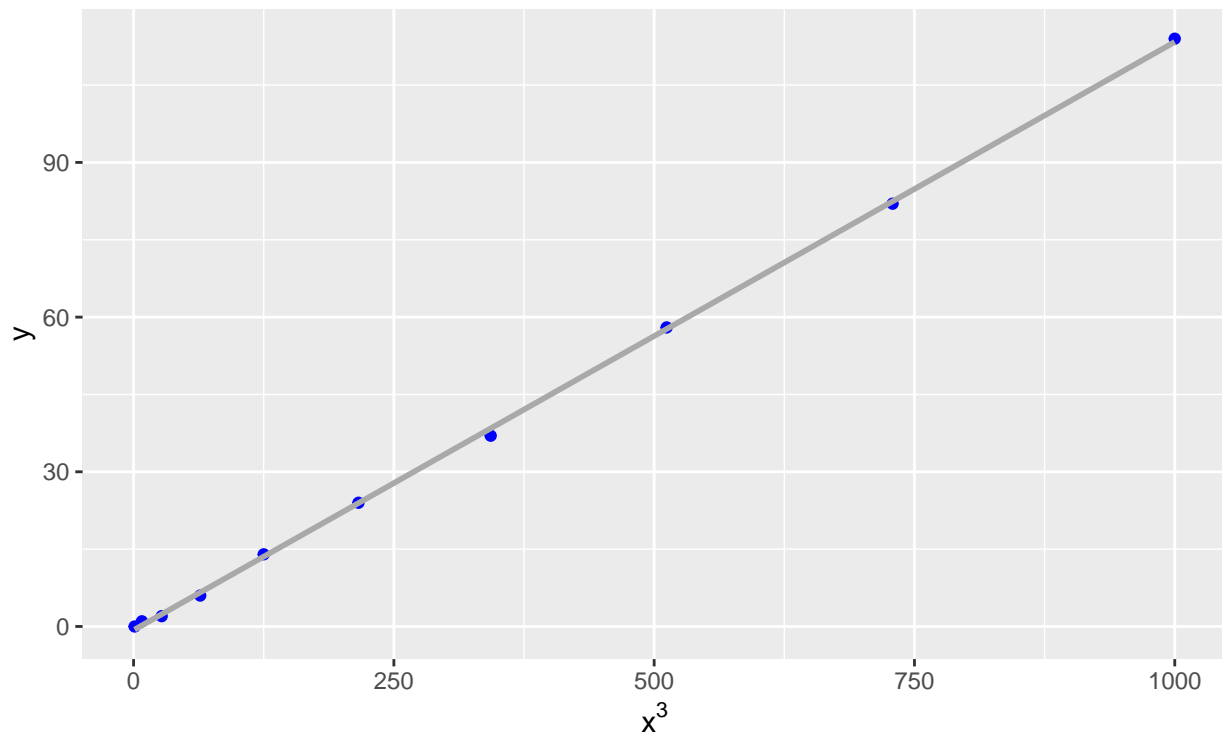
We can determine whether the proportionality model is reasonable by y values against x^3 . If the model is a good fit, then a fitted straight line passing through the origin should provide a reasonable approximation.

```
mydf3$x3 = x^3
```

```
mylm2 <- lm(y ~ x3, data = mydf3)
```

```
g <- ggplot(mydf3, aes(x^3, y)) + geom_point(col = "blue")
g <- g + xlab(TeX("$x^3$")) + ggtitle(TeX("Plot of y vs. $x^3$"))
g <- g + labs(subtitle = "with fitted OLS Line")
g + geom_smooth(method = "lm", color = "darkgray", se = FALSE)
```

Plot of y vs. x^3
with fitted OLS Line



```
summary(mylm2)
```

```
##
## Call:
## lm(formula = y ~ x3, data = mydf3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4200 -0.4326  0.1844  0.5571  0.7949
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.7074611   0.3196464  -2.213   0.0578 .
## x3           0.1140743   0.0007186 158.736 2.78e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7411 on 8 degrees of freedom
## Multiple R-squared:  0.9997, Adjusted R-squared:  0.9996
## F-statistic: 2.52e+04 on 1 and 8 DF, p-value: 2.775e-15
```

The proposed proportionality model appears to provide a good approximation:

-The R^2 statistic for our fitted, OLS model is 0.9997, which is indicative of a strong model fit.

-While the model's estimate of the y-intercept is not zero, the relatively high p-value for our estimate suggests that we do not have compelling evidence to reject the null hypothesis of a zero value for the intercept.

Tests exist to measure the percentage of body fat. Assume that such tests are accurate and that a great many carefully collected data are available. You may specify any other statistics, such as waist size and height, that you would like collected. Explain how the data could be arranged to check the assumptions underlying the submodels in this section. For example, suppose the data for males between ages 17 and 21 with constant body fat and height are examined. Explain how the assumption of constant density of the inner core could be checked.

The inner core is composed of muscle and bone, each of which has a different density. In order to find the percentage of the total volume of the inner core occupied by the bones. It is necessary to find whether the bones are proportional to the cube of the height. Some people have strong bones and others may not. Hence we can group these people with stronger bones into one category and those with weaker bones into another category. This helps to study the nature and assumptions that to be set for the inner core.