STAT 3355 - HW 8

Yebom Kim

2024-04-25

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

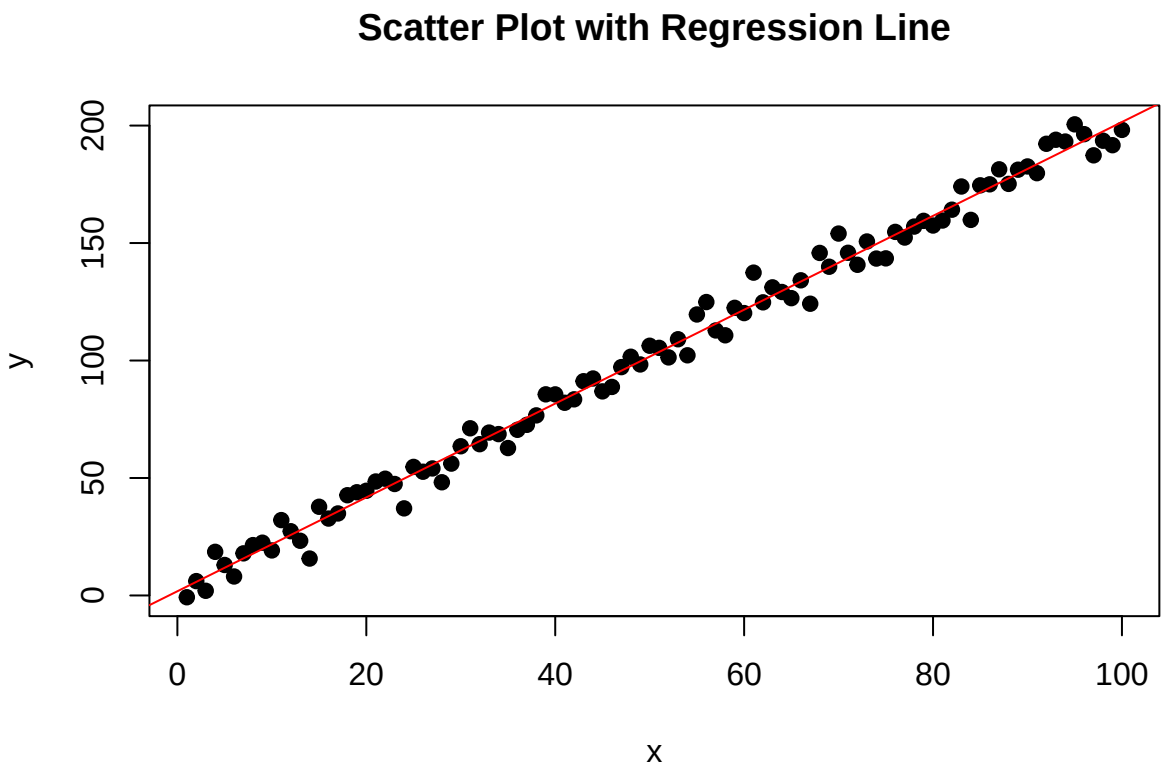
```
##
## Attaching package: 'dplyr'
```

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

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

Problem 1

```
set.seed(1)
x <- 1:100
epsilon <- rnorm(100, mean = 0, sd = 6)
y <- 1 + 2*x + epsilon

plot(x, y, main = "Scatter Plot with Regression Line", xlab = "x", ylab = "y", pch = 19)

model <- lm(y ~ x)
abline(model, col = "red")
```




```r
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.0403  -3.6350   0.0931   3.5109  13.7848
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.78999    1.09138    1.64    0.104
## x            1.99729    0.01876  106.45   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.416 on 98 degrees of freedom
## Multiple R-squared:  0.9914, Adjusted R-squared:  0.9913
## F-statistic: 1.133e+04 on 1 and 98 DF,  p-value: < 2.2e-16
```

```r
p_value <- summary(model)$coefficients[2,4]
```

```r
set.seed(1)
x <- 1:100
epsilon <- rnorm(100, mean = 0, sd = 6)
Y <- 1 + 2*x + epsilon
model <- lm(Y ~ x)
summary(model)
```

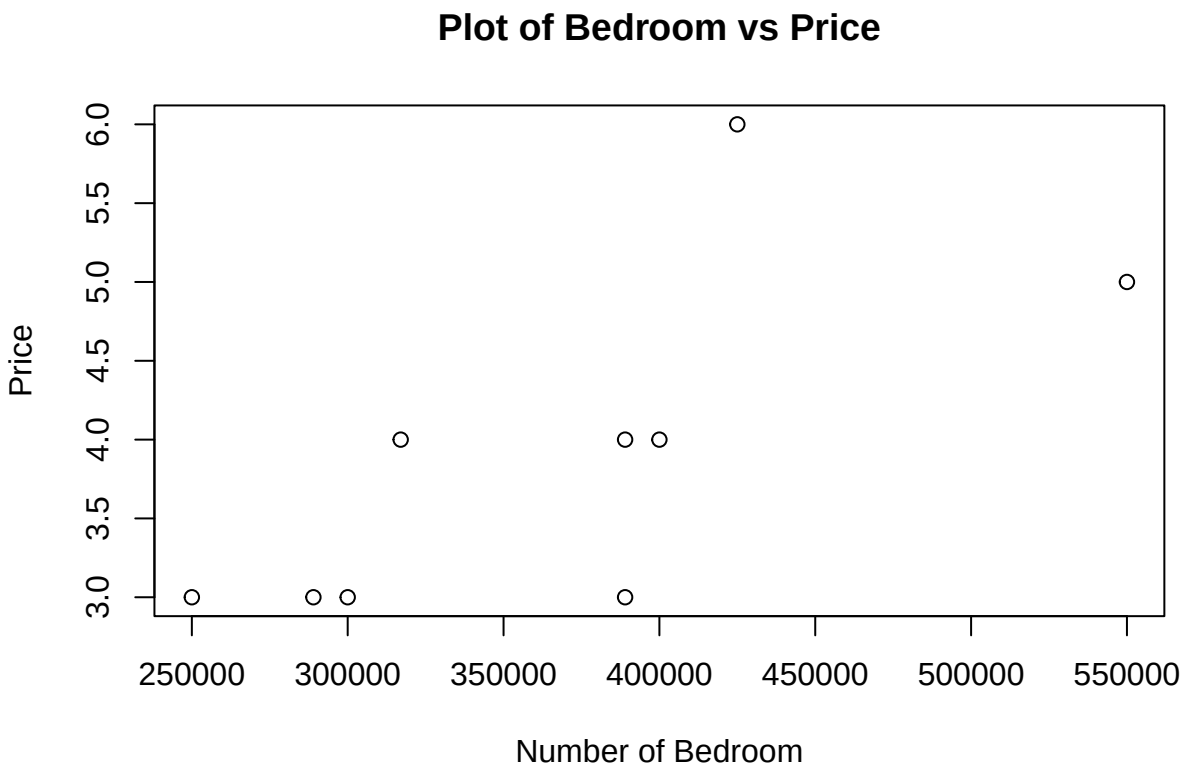
```
##
## Call:
## lm(formula = Y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.0403  -3.6350   0.0931   3.5109  13.7848
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.78999    1.09138    1.64    0.104
## x            1.99729    0.01876  106.45   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.416 on 98 degrees of freedom
## Multiple R-squared:  0.9914, Adjusted R-squared:  0.9913
## F-statistic: 1.133e+04 on 1 and 98 DF,  p-value: < 2.2e-16
```

```r
beta1 <- coef(model)[2]
se_beta1 <- summary(model)$coefficients[2, 2]

# Compute the t-value
t_value <- beta1 / se_beta1

# Compute the p-value
p_value <- 2 * pt(abs(t_value), df = length(Y) - 2)

# Compare p-value to significance level
alpha <- 0.05
if (p_value < alpha) {
  print("Reject the null hypothesis: beta1 is not equal to 2")
} else {
  print("Fail to reject the null hypothesis: beta1 is equal to 2")
}
```

```
## [1] "Fail to reject the null hypothesis: beta1 is equal to 2"
```

Problem 2

```r
price <- c(300000, 250000, 400000, 550000, 317000, 389000, 425000, 289000, 389000)
num_of_room <- c(3, 3, 4, 5, 4, 3, 6, 3, 4)
plot(price, num_of_room, main = "Plot of Bedroom vs Price", xlab = "Number of Bedroom", ylab = "Price")
```




```
predicted <- 2:8
predicted_prices <- predict(model, newdata = data.frame(bedrooms = predicted), interval = "confidence")
```

```
## Warning: 'newdata' had 7 rows but variables found have 100 rows
```

```
cat("\nConfidence intervals for mean price of houses with 2 to 8 bedrooms:\n")
```

```
##
## Confidence intervals for mean price of houses with 2 to 8 bedrooms:
```

```
for (i in 1:nrow(predicted_prices)) {
  cat("Bedrooms:", predicted[i], " - ", "CI:", predicted_prices[i, 2], "to", predicted_prices[i, 3], "\n")
}
```

```
## Bedrooms: 2  -  CI: 1.653716 to 5.92086
## Bedrooms: 3  -  CI: 3.68309 to 7.886073
## Bedrooms: 4  -  CI: 5.712292 to 9.851458
## Bedrooms: 5  -  CI: 7.741313 to 11.81702
## Bedrooms: 6  -  CI: 9.770145 to 13.78278
## Bedrooms: 7  -  CI: 11.79878 to 15.74873
## Bedrooms: 8  -  CI: 13.8272 to 17.7149
## Bedrooms: NA  -  CI: 15.85541 to 19.68128
## Bedrooms: NA  -  CI: 17.88339 to 21.64788
## Bedrooms: NA  -  CI: 19.91113 to 23.61473
```

```
## Bedrooms: NA  -  CI: 21.93861 to 25.58184
## Bedrooms: NA  -  CI: 23.96584 to 27.5492
## Bedrooms: NA  -  CI: 25.99278 to 29.51685
## Bedrooms: NA  -  CI: 28.01943 to 31.48478
## Bedrooms: NA  -  CI: 30.04577 to 33.45303
## Bedrooms: NA  -  CI: 32.07179 to 35.4216
## Bedrooms: NA  -  CI: 34.09746 to 37.39051
## Bedrooms: NA  -  CI: 36.12278 to 39.35978
## Bedrooms: NA  -  CI: 38.14772 to 41.32943
## Bedrooms: NA  -  CI: 40.17226 to 43.29947
## Bedrooms: NA  -  CI: 42.19639 to 45.26993
## Bedrooms: NA  -  CI: 44.22007 to 47.24084
## Bedrooms: NA  -  CI: 46.24329 to 49.21221
## Bedrooms: NA  -  CI: 48.26601 to 51.18407
## Bedrooms: NA  -  CI: 50.28823 to 53.15645
## Bedrooms: NA  -  CI: 52.30989 to 55.12937
## Bedrooms: NA  -  CI: 54.33099 to 57.10286
## Bedrooms: NA  -  CI: 56.35149 to 59.07695
## Bedrooms: NA  -  CI: 58.37135 to 61.05167
## Bedrooms: NA  -  CI: 60.39054 to 63.02706
## Bedrooms: NA  -  CI: 62.40904 to 65.00315
## Bedrooms: NA  -  CI: 64.42681 to 66.97998
## Bedrooms: NA  -  CI: 66.4438 to 68.95757
## Bedrooms: NA  -  CI: 68.45999 to 70.93597
## Bedrooms: NA  -  CI: 70.47533 to 72.91522
## Bedrooms: NA  -  CI: 72.48979 to 74.89534
## Bedrooms: NA  -  CI: 74.50333 to 76.87639
## Bedrooms: NA  -  CI: 76.51591 to 78.8584
## Bedrooms: NA  -  CI: 78.5275 to 80.8414
## Bedrooms: NA  -  CI: 80.53805 to 82.82543
## Bedrooms: NA  -  CI: 82.54753 to 84.81054
## Bedrooms: NA  -  CI: 84.5559 to 86.79675
## Bedrooms: NA  -  CI: 86.56314 to 88.7841
## Bedrooms: NA  -  CI: 88.56921 to 90.77262
## Bedrooms: NA  -  CI: 90.57408 to 92.76234
## Bedrooms: NA  -  CI: 92.57772 to 94.75328
## Bedrooms: NA  -  CI: 94.58013 to 96.74547
## Bedrooms: NA  -  CI: 96.58127 to 98.73891
## Bedrooms: NA  -  CI: 98.58114 to 100.7336
## Bedrooms: NA  -  CI: 100.5797 to 102.7296
## Bedrooms: NA  -  CI: 102.577 to 104.7269
## Bedrooms: NA  -  CI: 104.573 to 106.7255
## Bedrooms: NA  -  CI: 106.5677 to 108.7254
## Bedrooms: NA  -  CI: 108.5612 to 110.7265
## Bedrooms: NA  -  CI: 110.5534 to 112.7289
## Bedrooms: NA  -  CI: 112.5443 to 114.7326
## Bedrooms: NA  -  CI: 114.534 to 116.7374
## Bedrooms: NA  -  CI: 116.5225 to 118.7435
## Bedrooms: NA  -  CI: 118.5099 to 120.7507
## Bedrooms: NA  -  CI: 120.4961 to 122.7591
## Bedrooms: NA  -  CI: 122.4812 to 124.7686
## Bedrooms: NA  -  CI: 124.4653 to 126.7792
## Bedrooms: NA  -  CI: 126.4483 to 128.7907
## Bedrooms: NA  -  CI: 128.4303 to 130.8033
```

```
## Bedrooms: NA  -  CI: 130.4113 to 132.8169
## Bedrooms: NA  -  CI: 132.3914 to 134.8313
## Bedrooms: NA  -  CI: 134.3707 to 136.8467
## Bedrooms: NA  -  CI: 136.3491 to 138.8628
## Bedrooms: NA  -  CI: 138.3267 to 140.8798
## Bedrooms: NA  -  CI: 140.3035 to 142.8976
## Bedrooms: NA  -  CI: 142.2796 to 144.9161
## Bedrooms: NA  -  CI: 144.255 to 146.9353
## Bedrooms: NA  -  CI: 146.2297 to 148.9552
## Bedrooms: NA  -  CI: 148.2038 to 150.9757
## Bedrooms: NA  -  CI: 150.1773 to 152.9968
## Bedrooms: NA  -  CI: 152.1502 to 155.0184
## Bedrooms: NA  -  CI: 154.1226 to 157.0406
## Bedrooms: NA  -  CI: 156.0944 to 159.0634
## Bedrooms: NA  -  CI: 158.0658 to 161.0866
## Bedrooms: NA  -  CI: 160.0367 to 163.1103
## Bedrooms: NA  -  CI: 162.0072 to 165.1344
## Bedrooms: NA  -  CI: 163.9772 to 167.1589
## Bedrooms: NA  -  CI: 165.9469 to 169.1839
## Bedrooms: NA  -  CI: 167.9161 to 171.2092
## Bedrooms: NA  -  CI: 169.885 to 173.2349
## Bedrooms: NA  -  CI: 171.8536 to 175.2609
## Bedrooms: NA  -  CI: 173.8219 to 177.2872
## Bedrooms: NA  -  CI: 175.7898 to 179.3139
## Bedrooms: NA  -  CI: 177.7574 to 181.3408
## Bedrooms: NA  -  CI: 179.7248 to 183.368
## Bedrooms: NA  -  CI: 181.6919 to 185.3955
## Bedrooms: NA  -  CI: 183.6588 to 187.4233
## Bedrooms: NA  -  CI: 185.6254 to 189.4512
## Bedrooms: NA  -  CI: 187.5918 to 191.4794
## Bedrooms: NA  -  CI: 189.5579 to 193.5079
## Bedrooms: NA  -  CI: 191.5239 to 195.5365
## Bedrooms: NA  -  CI: 193.4896 to 197.5653
## Bedrooms: NA  -  CI: 195.4552 to 199.5944
## Bedrooms: NA  -  CI: 197.4206 to 201.6236
## Bedrooms: NA  -  CI: 199.3858 to 203.6529
```

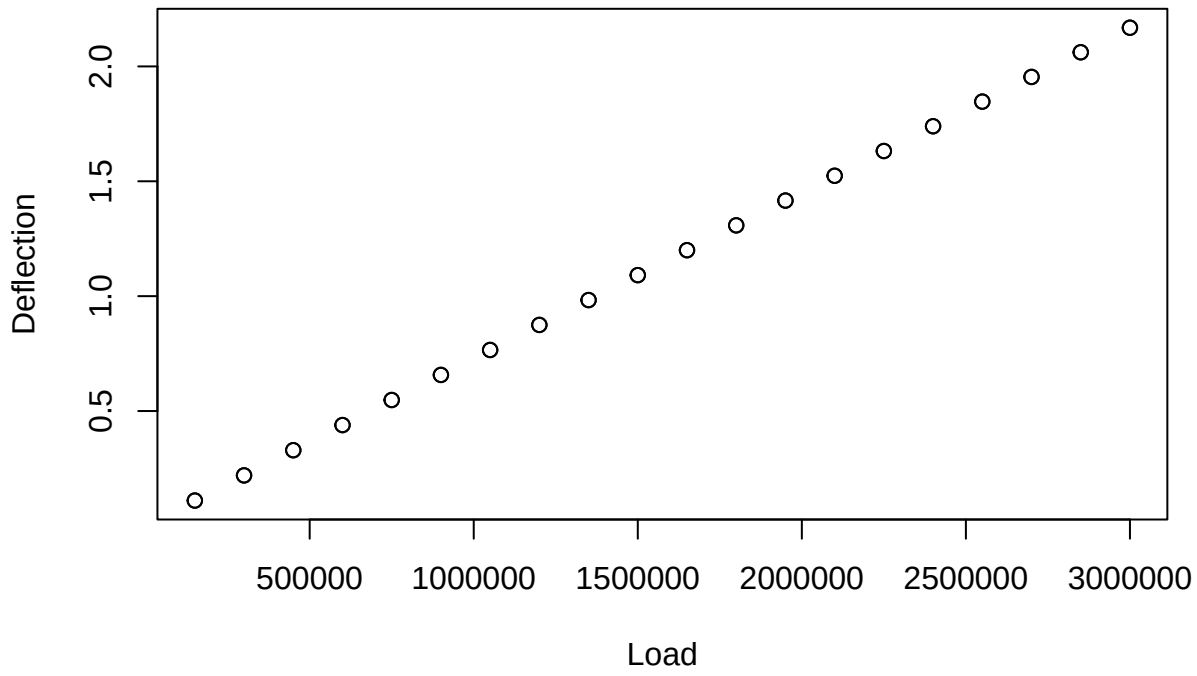
Problem 3

```
library(UsingR)
```

```
## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

## Loading required package: HistData
```

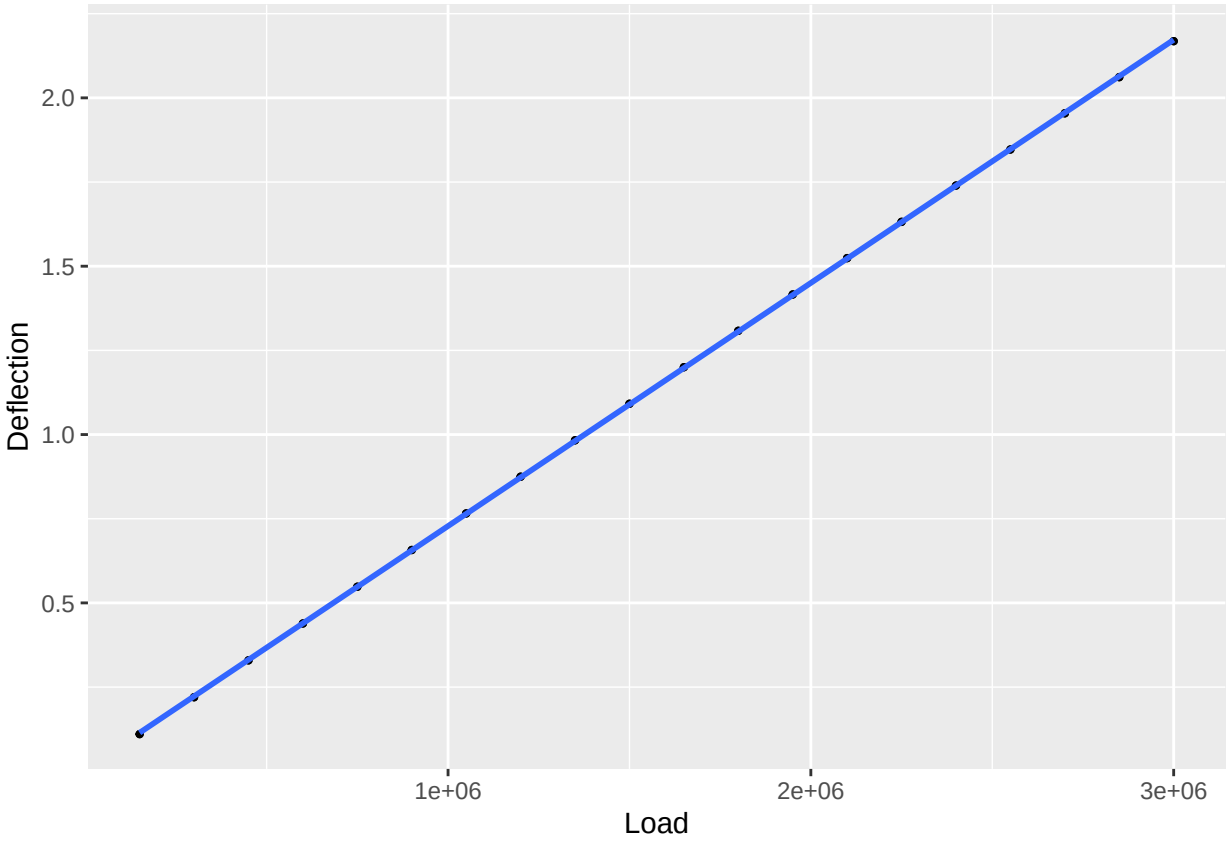
```
## Loading required package: Hmisc

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
data("deflection")
data = data.frame(deflection)
n=length(data$Deflection)
attach(data)
fit=lm(Deflection~Load)
b0=as.numeric(fit$coefficients[1])
b1=as.numeric(fit$coefficients[2])
plot(Load,Deflection)
```




```
ggplot(data , aes(x = Load, y = Deflection)) + geom_point(size = 0.8) + geom_smooth(method = lm)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```


2.0
1.5
1.0
0.5
Deflection
1e+06
2e+06
3e+06
Load