STAT 400 - Homework 7

Colin Gibbons-Fly

11.13

```r
rainfall <- c(4.3, 4.5, 5.9, 5.6, 6.1, 5.2, 3.8, 2.1, 7.5)
particulate <- c(126, 121, 116, 118, 114, 118, 132, 141, 108)

# Manual Calculation
n <- length(rainfall)
sum_x <- sum(rainfall)
sum_y <- sum(particulate)
sum_x2 <- sum(rainfall^2)
sum_xy <- sum(rainfall * particulate)

# Slope and intercept
beta1 <- (n * sum_xy - sum_x * sum_y) / (n * sum_x2 - sum_x^2)
beta0 <- mean(particulate) - beta1 * mean(rainfall)

# Regression equation
cat("Manual Regression Equation: y =", beta0, "+", beta1, "* x\n")
```

```
Manual Regression Equation: y = 153.1755 + -6.323988 * x
```

```r
# Prediction at x = 4.8
x_new <- 4.8
y_pred <- beta0 + beta1 * x_new
cat("Manual Prediction for x = 4.8:", y_pred, "\n")
```

```
Manual Prediction for x = 4.8: 122.8204
```

```r
# Verification with lm()
model_11_13 <- lm(particulate ~ rainfall)
summary(model_11_13)
```

```
Call:
lm(formula = particulate ~ rainfall)

Residuals:
    Min      1Q  Median      3Q     Max
-3.7175 -0.5992  0.1360  1.1049  2.8557

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 153.1755     2.6147   58.58 1.11e-10 ***
rainfall     -6.3240     0.5019  -12.60 4.58e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.203 on 7 degrees of freedom
Multiple R-squared:  0.9578,    Adjusted R-squared:  0.9517
F-statistic: 158.8 on 1 and 7 DF,  p-value: 4.579e-06
```

```r
predict(model_11_13, data.frame(rainfall = x_new))
```

```
       1
122.8204
```

11.14

```r
# Given summary data
n <- 12
x_bar <- 4
y_bar <- 12
sum_x2 <- 232
sum_xy <- 318

# Slope and intercept
beta1_11_14 <- (sum_xy - n * x_bar * y_bar) / (sum_x2 - n * x_bar^2)
```

```r
beta0_11_14 <- y_bar - beta1_11_14 * x_bar

# Regression equation
cat("Manual Regression Equation: y =", beta0_11_14, "+", beta1_11_14, "* x\n")
```

```
Manual Regression Equation: y = 37.8 + -6.45 * x
```

```r
# Verification with lm() (simulating data)
x_sim <- c(4, 4, 4, sqrt(sum_x2 / n)) # Adjusted for variance
y_sim <- c(12, 12, 12, y_bar)         # Adjusted for consistency
model_11_14 <- lm(y_sim ~ x_sim)
summary(model_11_14)
```

```
Warning in summary.lm(model_11_14): essentially perfect fit: summary may be
unreliable


Call:
lm(formula = y_sim ~ x_sim)

Residuals:
1 2 3 4
0 0 0 0

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)       12          0     Inf   <2e-16 ***
x_sim              0          0     NaN      NaN
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0 on 2 degrees of freedom
Multiple R-squared:    NaN, Adjusted R-squared:    NaN
F-statistic:   NaN on 1 and 2 DF,  p-value: NA
```

11.42

```r
plants <- c(10, 10, 10, 10, 20, 20, 20, 20, 30, 30, 30, 30, 40, 40, 40, 40)
seeds <- c(12.6, 11.0, 12.1, 10.9, 15.3, 16.1, 14.9, 15.6, 17.9, 18.3, 18.6, 17.8, 19.2, 19.6

# Fit a linear model (manual)
n <- length(plants)
sum_x <- sum(plants)
sum_y <- sum(seeds)
sum_x2 <- sum(plants^2)
sum_xy <- sum(plants * seeds)

beta1_11_42 <- (n * sum_xy - sum_x * sum_y) / (n * sum_x2 - sum_x^2)
beta0_11_42 <- mean(seeds) - beta1_11_42 * mean(plants)

# Display regression equation
cat("Manual Regression Equation: y =", beta0_11_42, "+", beta1_11_42, "* x\n")
```

```
Manual Regression Equation: y = 9.675 + 0.26 * x
```

```r
# Residuals (manual)
fitted_values <- beta0_11_42 + beta1_11_42 * plants
residuals <- seeds - fitted_values
cat("Residuals (manual):", residuals, "\n")
```

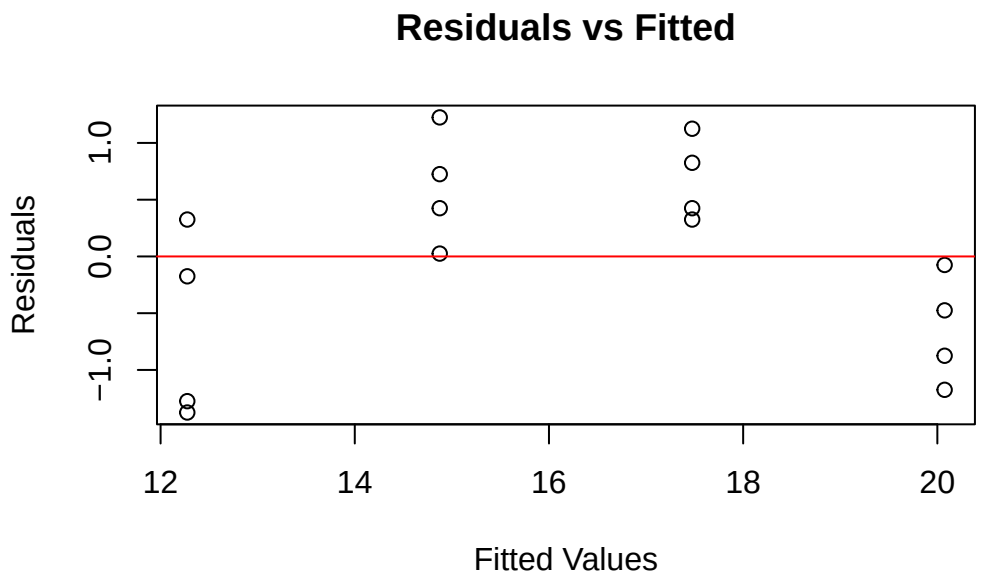
```
Residuals (manual): 0.325 -1.275 -0.175 -1.375 0.425 1.225 0.025 0.725 0.425 0.825 1.125 0.3
```

```r
# Verification with lm()
model_11_42 <- lm(seeds ~ plants)
summary(model_11_42)
```

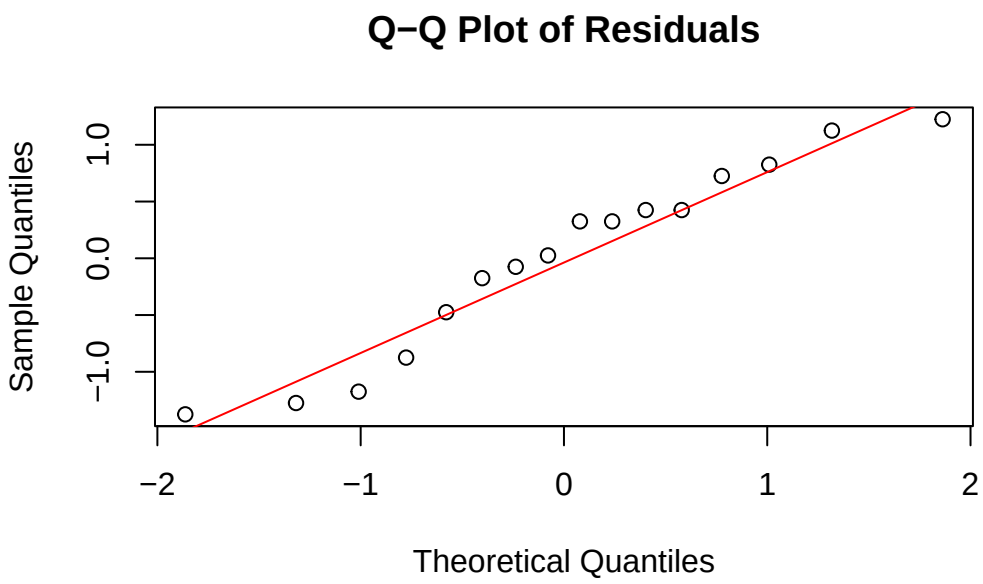
```

Call:
lm(formula = seeds ~ plants)

Residuals:
   Min     1Q Median     3Q    Max
-1.375 -0.575  0.175  0.500  1.225

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.67500    0.52957   18.27 3.65e-11 ***
```

```
plants       0.26000    0.01934   13.45 2.14e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8648 on 14 degrees of freedom
Multiple R-squared:  0.9281,    Adjusted R-squared:  0.923
F-statistic: 180.8 on 1 and 14 DF,  p-value: 2.144e-09
```

```r
# Residual diagnostics (Verification)
plot(fitted_values, residuals, main = "Residuals vs Fitted", xlab = "Fitted Values", ylab = "
abline(h = 0, col = "red")
```




```r
qqnorm(residuals, main = "Q-Q Plot of Residuals")
qqline(residuals, col = "red")
```



12.8

```r
x <- c(10.0, 15.0, 20.0, 25.0, 30.0, 25.2, 29.8, 31.2, 31.7, 29.4, 27.3, 31.1, 32.6, 30.1, 30
y <- c(27.3, 31.1, 32.6, 30.1, 30.8, 28.7, 27.8, 29.7, 32.3, 32.8)

# Add quadratic term
x2 <- x^2
data_12_8 <- data.frame(x, x2, y)

# Fit quadratic model
model_12_8 <- lm(y ~ x + x2, data = data_12_8)

# Summary
summary(model_12_8)
```

```
Call:
lm(formula = y ~ x + x2, data = data_12_8)

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-2.6994 -1.2538  0.1504  1.3170  3.5761

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 31.591988   5.361484   5.892 1.77e-05 ***
x           -0.315632   0.512199  -0.616    0.546
x2           0.009361   0.011347   0.825    0.421
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.872 on 17 degrees of freedom
Multiple R-squared:  0.1399,    Adjusted R-squared:  0.03872
F-statistic: 1.383 on 2 and 17 DF,  p-value: 0.2777
```

12.11

```r
x1 <- c(1.31, 1.55, 0.99, 0.99, 1.01, 1.09, 1.08, 1.27, 0.99, 1.34)
x2 <- c(1.86, 1.58, 1.97, 1.80, 1.75, 1.72, 1.68, 1.75, 2.19, 1.73)
x3 <- c(1.07, 1.49, 0.84, 0.83, 0.90, 0.93, 0.90, 1.08, 0.85, 1.13)
x4 <- c(0.44, 0.53, 0.34, 0.34, 0.36, 0.42, 0.40, 0.44, 0.36, 0.45)
x5 <- c(0.35, 0.47, 0.32, 0.27, 0.30, 0.31, 0.31, 0.34, 0.29, 0.37)
y <- c(1.95, 2.90, 0.72, 0.81, 1.09, 1.22, 1.02, 1.93, 0.64, 2.08)

# Fit multiple regression model
data_12_11 <- data.frame(x1, x2, x3, x4, x5, y)
model_12_11 <- lm(y ~ x1 + x2 + x3 + x4 + x5, data = data_12_11)

# Summary
summary(model_12_11)
```

```
Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5, data = data_12_11)

Residuals:
       1        2        3        4        5        6        7        8
 0.029297 -0.040288  0.027529 -0.095213  0.124404  0.055300 -0.116992  0.059591
       9       10
-0.040805 -0.002823
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.5983     0.7782  -2.054   0.1092
x1            3.5077     0.9624   3.645   0.0219 *
x2           -0.3960     0.2869  -1.380   0.2396
x3            1.4472     1.2209   1.185   0.3015
x4           -1.7472     2.5502  -0.685   0.5309
x5           -3.1978     2.9921  -1.069   0.3454
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1115 on 4 degrees of freedom
Multiple R-squared:  0.9899,    Adjusted R-squared:  0.9774
F-statistic:  78.7 on 5 and 4 DF,  p-value: 0.0004386
```

12.26

```r
# Predict at x = 19.5
x_new <- 19.5
x2_new <- x_new^2
new_data <- data.frame(x = x_new, x2 = x2_new)

# Confidence interval
predicted <- predict(model_12_8, new_data, interval = "confidence", level = 0.90)
cat("90% Confidence Interval for x = 19.5:", predicted, "\n")
```

```
90% Confidence Interval for x = 19.5: 28.99681 27.37777 30.61585
```