Homework 8

Brady Lamson

Section 12.1

Problem 7

- \$y = \$ 28-day standard cured strength (psi)
- \$x = \$ accelerated strength (psi)

$$y = 1800 + 1.3x$$

Part A

What is the expected value of 28-day strength when accelerated strength is 2500?

```r
x <- 2500

1800 + (1.3*x)
```

```
[1] 5050
```

Part B

By how much can we expect 28-day strength to change when accelerated strength increases by 1 psi?

The slope, so we can expect it to increase by 1.3 psi.

Part C

By how much can we expect 28-day strength to change when accelerated strength increases by 100 psi?

+130 psi

Part D

By how much can we expect 28-day strength to change when accelerated strength increases by 100 psi?

-130

Section 12.2

Problem 15

Part B

Is the value of strength completely and uniquely determined by the value of MOE? Explain.

I would wager no. If that was the case there would be no up and down fluctuation in strength as MOE increases. Obviously other things have an effect as well.

Part C

$$y = 3.2925 + 0.10748x$$

Where y is the strength and x is the modus of elasticity

```r
x <- 40
y <- 3.2925 + (0.10748*x)
y
```

```
[1] 7.5917
```

I wouldn't feel comfortable predicting a value for strength off of a MOE of 100. As we have no values that high we can't determine this relationship continues linearly.

Part D

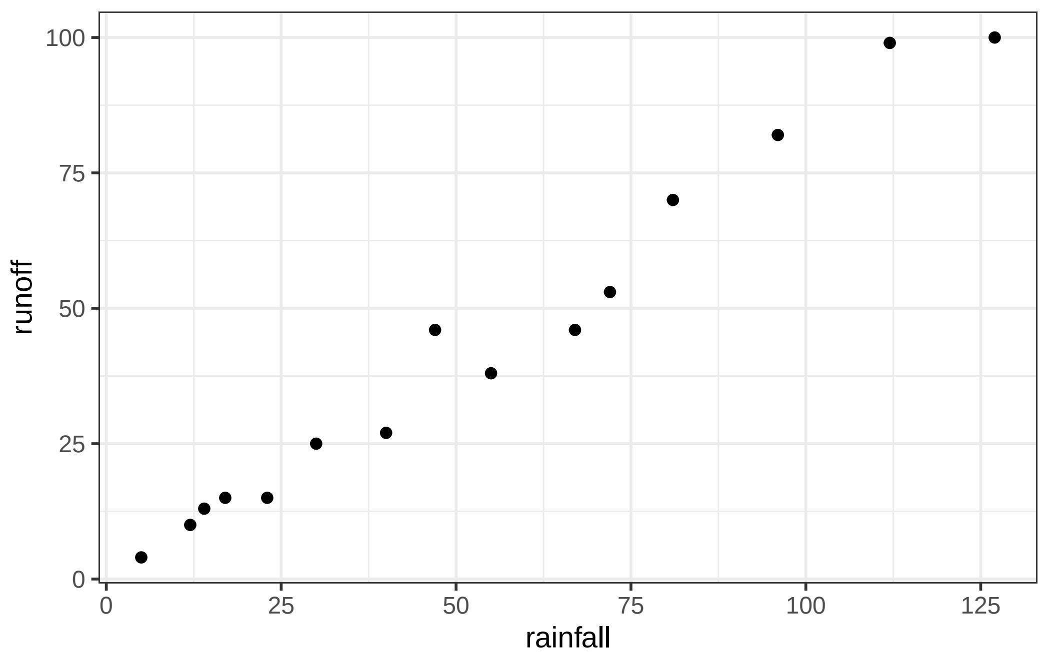
$$SSE = 18.736$$
$$SST = 71.605$$
$$r^2 = 0.738$$

They do indicate that the regression does a good job. If more of the total sum of squares was made up by the error and the r^2 was lower I would not say this.

Problem 16

```
df <- tibble(
  rainfall = c(5,12,14,17,23,30,40,47,55,67,72,81,96,112,127),
  runoff = c(4,10,13,15,15,25,27,46,38,46,53,70,82,99,100)
)

df %>%
  ggplot(aes(x=rainfall, y=runoff)) +
  geom_point() +
  theme_bw()
```



The scatterplot does support the use of a simple linear regression model.

Part B

```r
reg <- lm(runoff ~ rainfall, data = df)
reg %>% summary()
```

```
Call:
lm(formula = runoff ~ rainfall, data = df)

Residuals:
   Min     1Q Median     3Q    Max 
-8.279 -4.424  1.205  3.145  8.261 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.12830    2.36778  -0.477    0.642    
rainfall     0.82697    0.03652  22.642  7.9e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.24 on 13 degrees of freedom
Multiple R-squared:  0.9753,    Adjusted R-squared:  0.9734 
F-statistic: 512.7 on 1 and 13 DF,  p-value: 7.896e-12
```

Point estimate of the slope is 0.82697. Point estimate of the intercept is -1.12830.

Part C

```r
predict.lm(reg, newdata = data.frame(rainfall = c(50)))
```

```
       1 
40.22035 
```

Part D

```r
reg %>% anova()
```

```
Analysis of Variance Table

Response: runoff
          Df Sum Sq Mean Sq F value    Pr(>F)
rainfall   1  14079 14078.7  512.65 7.896e-12 ***
Residuals 13    357    27.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$\sigma = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{27.5}{13}} \approx 1.45$$

Part E

0.9753

Problem 17

Part A

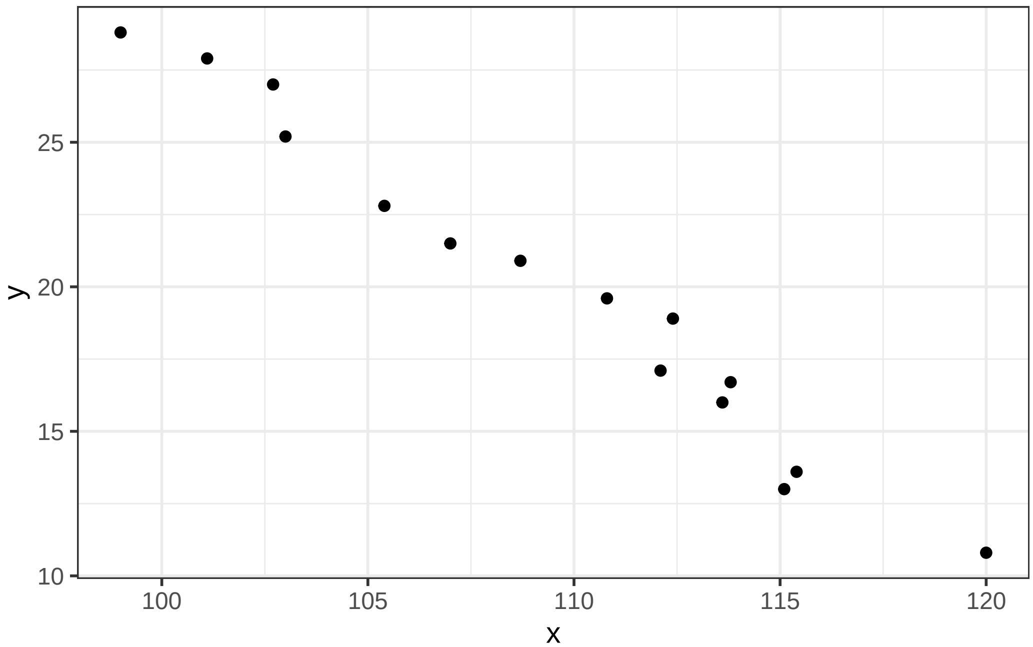
```
Call:
lm(formula = y ~ x, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-1.7754 -0.5727 -0.1325  0.6034  1.6818

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 118.90992    4.49912   26.43 1.10e-12 ***
x            -0.90473    0.04109  -22.02 1.12e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.938 on 13 degrees of freedom
Multiple R-squared:  0.9739,    Adjusted R-squared:  0.9719
F-statistic: 484.8 on 1 and 13 DF,  p-value: 1.125e-11
```

$$y = 118.91 - 0.90x$$



Based on the output of the linear regression and the scatterplot, it does appear that the model is going to explain a great deal of the observed variation in y.

Part B

The slope indicates that as x increases, y decreases.

Part C

```r
x <- 135
y <- 118.91 - (.9*x)
y
```

```
[1] -2.59
```

You get a negative value for porosity percent which should be impossible. You shouldn't predict beyond the max or minimum x values as you can't assume the linear relationship will continue forever.

Part D

Below are three ways to get the residuals.

```r
reg$residuals[1:2]
```

```
         1          2 
-0.5415817  0.4583527 
```

```r
df$y[1:2] - predict(reg, data.frame(x=c(99,101.1)))
```

```
         1          2 
-0.5415817  0.4583527 
```

```r
df$y[1:2] - (118.91 - (.90473*df$x[1:2]))
```

```
[1] -0.541730  0.458203
```

Part E

```
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)    
x          1 426.62  426.62  484.84 1.125e-11 ***
Residuals 13  11.44    0.88                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$\sigma = \sqrt{\frac{11.44}{13}} \approx 0.938$$

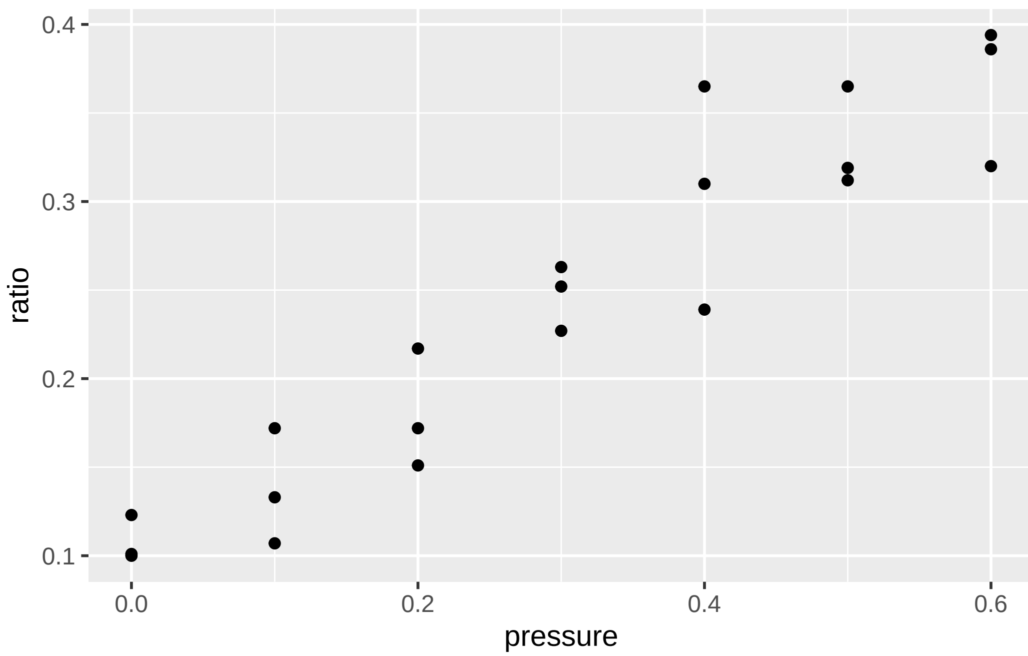
Part F

What proportion of observed variation in porosity can be attributed linear relationship between unit weight and porosity?

0.9739

Problem 20

Part A



A scatter plot does support the use of a simple linear regression model.

Part B

Slope 0.461, intercept = 0.101

Part C

```
intercept <- 0.101
slope <- 0.461
x <- 0.45
y <- intercept + (slope * x)
y
```

```
[1] 0.30845
```

Part D

```
MSE <- 0.00110
sd <- sqrt(MSE)

sd
```

```
[1] 0.03316625
```

On average, the difference between the predicted value and the real value of this data is approximately 0.03.

Part E

The total variation is 0.19929 and 0.895 of it is explained by the linear model.

Section 12.3

Problem 32

Part A

Based on the minitab output there is a useful relationship between rainfall and runoff. I determined this using the p-value of the slope (rainfall) being less than 0.05 and the r^2 value of 0.975 which indicates the model describes the vast majority of the variance in the data.

To create a confidence interval we use:

$$
\begin{aligned}
CI &= \widehat{\beta}_1 \pm (t_{\alpha/2,n-2} \cdot s_{\hat{\beta}_1}) \\
&= 0.82647 \pm (t_{0.05,13} \cdot 0.03652) \\
&= 0.82647 \pm (1.77 \cdot 0.03652) \\
CI &= [0.762, 0.892]
\end{aligned}
$$

```r
beta_hat_1 <- 0.82697
t <- qt(0.95, 13, lower.tail = T)
s_beta <- 0.03652

(beta_hat_1 + (c(-t,t) * s_beta)) %>% round(3)
```

```
[1] 0.762 0.892
```

Problem 33

Calculate a 95% CI for the slope of the population regression line.

$$
\begin{aligned}
n &= 26 + 1 = 27 \\
\widehat{\beta}_1 &\approx 0.10748 \\
t_{0.025,25} &= \pm 2.06 \\
s_{\hat{\beta}_1} &= 0.01280 \\
CI &= 0.10748 \pm (2.06 \cdot 0.01280) \\
CI &= [0.081, 0.134]
\end{aligned}
$$

For every GPa the MOE increases, the flexural strength experiences an increase between 0.081 and 0.134 MPa on average.

Part B

$$
\begin{aligned}
&H_0 : \beta_1 = 0.1 \\
&H_a : \beta_1 > 0.1 \\
&\alpha = 0.05 \\
&t = \frac{0.10748 - 0.1}{0.0128} \\
&t = 0.584 \\
&p = t_{cdf}(-2.4867, 18 - 2) \\
&p \approx 0.2882 \\
&p > \alpha
\end{aligned}
$$

As $p > \alpha$ we fail to reject the null hypothesis that the slope is equal to 0.1.

Problem 34

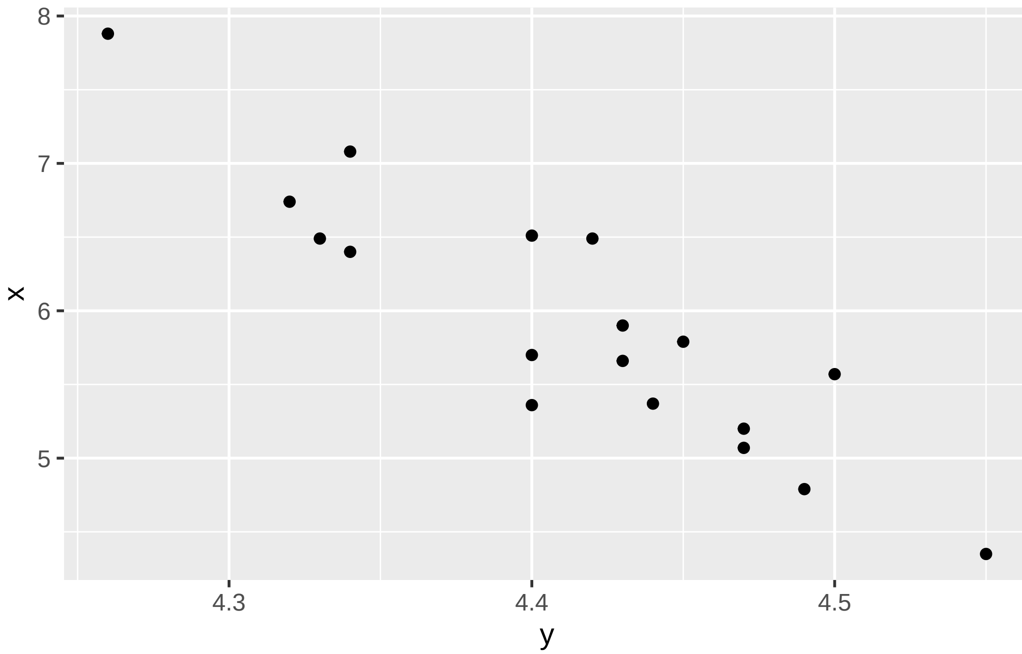
Part A

$$y = 4.859 + (-0.075 \cdot x)$$

Part B

$$r^2 = 0.766$$

Part C



Based on the r^2 and the scatterplot there appears to be a useful linear relationship.

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

We can utilize the p-value of the anova table for this problem. We have $F = 56.635$ and $p = 1.213e^{-06}$ so $p < 0.05$ so we can reject the null hypothesis that there is not a linear relationship between dielectric constant and air void.

Part D

$$
\begin{aligned}
&H_0 : \beta_1 = -0.05 \\
&H_a : \beta_1 \neq -0.05 \\
&\alpha = 0.01 \\
&t = \frac{-0.074676 - (-0.05)}{0.009923} \\
&t = -2.4867 \\
&p = t_{cdf}(-2.4867, 18 - 2) \\
&p \approx 0 \\
&p < \alpha
\end{aligned}
$$

As $p < \alpha$ we can reject the null hypothesis that the slope is -0.05.