STM4PSD – Workshop 12 Solutions

1. The output is as below.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  19.8669     2.0581   9.653   <2e-16 ***
Sun          10.2130     0.3086  33.092   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.83 on 320 degrees of freedom
  (27 observations deleted due to missingness)
Multiple R-squared:  0.7739,Adjusted R-squared:  0.7732
F-statistic:  1095 on 1 and 320 DF,  p-value: < 2.2e-16
```

   (a) We have, to three decimal places,

   $$\widehat{\text{Yield}} = 19.867 + 10.213 \times \text{Sun}.$$

   (b) The estimate is 10.21. We estimate that, on average, Yield will increase by 10.213 kWh for every 1 hour increase in bright sunshine.

   (c) Since the p-value for this test is approximately 0 (so less than 0.05), we reject the null hypothesis.

   (d) We have enough evidence to suggest that there is a significant linear association between yield and sunshine. That is, as sunshine hours increase so too do we expect the yield of the solar plant to increase.

   (e) Yes. We have $R^2 = 0.7739$ which suggests that more than 77% of the variation in yield can be explained by the estimated model.

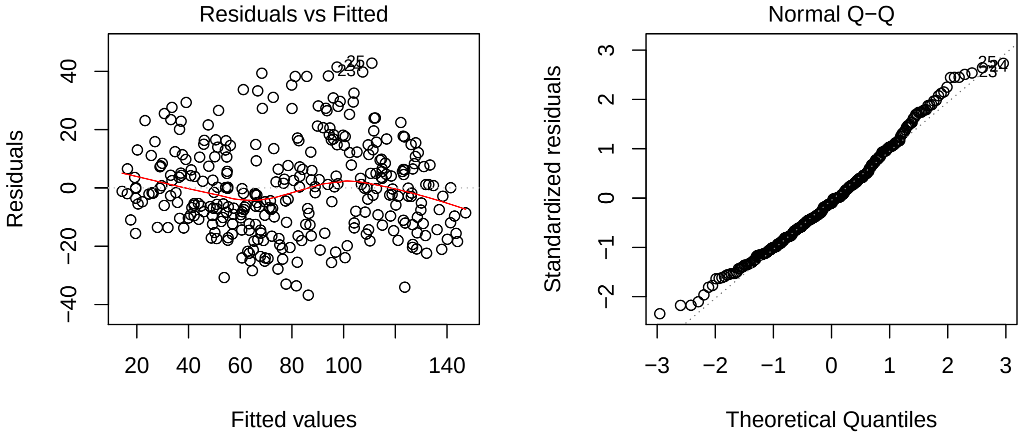
   (f) Using `qt(0.975, df = 320)` we obtain $t_{320,0.975} = 1.967$. Hence, our 95% confidence interval for β_1 is $10.213 \pm 1.967 \times 0.309 = (9.605,\ 10.821)$.

2. The output is below.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1.349      2.849  -0.474    0.636
Sun            8.558      0.323  26.494   <2e-16 ***
MaxTemp        1.496      0.154   9.713   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.73 on 317 degrees of freedom
  (29 observations deleted due to missingness)
Multiple R-squared:  0.8238,Adjusted R-squared:  0.8227
F-statistic: 740.9 on 2 and 317 DF,  p-value: < 2.2e-16
```

   (a) The residuals versus fits plot and the Q-Q plot of the residuals are below.



There are no patterns or 'fanning' in the residuals vs fits plot suggesting independence between residuals and fits and constant error variance. The Q-Q plot of residuals is approximately linear so there are no clear violations of normality.

(b) Yes, very well. The R^2 value has now increase to more than 0.82 suggesting a very good fit.

(c) The estimate is 1.496. We estimate that, on average, Yield will increase by 1.496 kWh for every 1 degree Celsius increase of maximum temperature.

(d) Since the p-value for this test is approximately 0, we reject the null hypothesis that $\beta_2 = 0$. We therefore have found evidence of a linear association between yield and maximum temperature.

(e) Using `qt(0.975, df = 317)` we obtain $t_{317,0.975} = 1.967$. Hence, our 95% confidence interval for β_2 is $1.496 \pm 1.967 \times 0.154 = (1.193,\ 1.799)$.

3. Continuing the previous question, we will now obtain confidence and prediction intervals for the response for a set value of the explanatory variables.

(a) The new data frame looks like:

```
  MaxTemp Sun
1      23   1
```

(b) We obtain the output below.

```
       fit      lwr      upr
1 41.61376 37.70476 45.52277
```

An example statement could be: For a maximum temperature of 23 degrees Celsius and 1 hour of bright sunshine, we are 95% confident that the average yield is between 37.7 and 45.5 kWh. The estimated mean yield is 41.6 kWh.

(c) We obtain:

```
       fit      lwr      upr
1 41.61376 10.42856 72.79896
```

An example statement could be: For a maximum temperature of 23 degrees Celsius and 1 hour of bright sunshine, we predict with 95% confidence that the resulting yield will be between 10.4 and 72.8 kWh.

(d) The confidence interval is an interval estimate for average yield whereas the prediction interval is an estimate for a single yield (i.e. yield on a single day). The prediction interval is wider to take into account the error variance.

(e) Yes, since the bounds for the intervals have increased substantially. For example, by increasing sunshine to 10 hours, we are now highly confident that the yield on any given day will be at least 87.6 kWh..