2. (a) Scatterplot:

The data sets given for analyzing the cherry trees contain two quantitative variables: the day in April and the March temperature. We could tell from the scatterplot that the association between the temperature in March and the day in April for the first cherry blossom is fairly strong, negative, moderately liner with no unusual features.

|  |
| --- |
| Coefficients:  (Intercept) Mar  33.120 -4.686 |

(b) The simple regression model is:

AprDays = -4.686 × MarTemp + 33.120



Residual Plot:

The model is appropriate because the residuals spread evenly rather than following into any shape.

More days taken in April for the first cherry blossom are associated with lower temperature in March. Higher temperature in March, less days (around 4 to 5 days for every 1 degree Celsius increase) are taken for cherry trees to blossom.

|  |
| --- |
| Residual standard error: 3.022 on 22 degrees of freedom  Multiple R-squared: 0.7243, Adjusted R-squared: 0.7118  F-statistic: 57.8 on 1 and 22 DF, p-value: 1.36e-07 |

(c) R²:

R²= 0.7243

R² shows that 72.43% of the variablilty in the AprDay is accounted for by variation in MarTemp, which implies that the model does a good job.

(d) LSRL:

The slope of LSRL is -4.686

(e) Normality:

The residuals do not seem to deviate from a random sample from a normal distribution in any systematic manner, so the assumption of normality is retained.

|  |
| --- |
| Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 33.1203 2.7333 12.117 3.30e-11 \*\*\*  Mar -4.6855 0.6163 -7.603 1.36e-07 \*\*\* |

(f)

For H0: β=0 vs. H1: β≠0

We could tell that p-value = 1.36e-07 is a relatively small value (< α=0.05). Base on that, we could reject the null hypothesis H0 and conclude that there is sufficient evidence that the slope is different than zero.

3.