Question 1:

a)

$$\frac{\partial(w_0)}{w_0} = \sum_{n=1}^{N} (y_n - w_0 - w_1 x_n)^2 = \sum_{n=1}^{N} 2(y_n - w_0 - w_1 x_n) \cdot (0 - 1 - 0)$$

$$\frac{\partial(w_0)}{w_0} = 2 \sum_{n=1}^{N} -y_n + w_0 + w_1 x_n$$

$$0 = -2 \sum_{n=1}^{N} y_n + 2w_0 \sum_{n=1}^{N} + 2w_1 \sum_{n=1}^{N} x_n$$

$$2 \sum_{n=1}^{N} y_n - 2w_1 \sum_{n=1}^{N} x_n = 2w_0 \sum_{n=1}^{N}$$

$$2\left(\sum_{n=1}^{N} y_n - w_1 \sum_{n=1}^{N} x_n\right) = 2w_0 N$$

$$\frac{\cancel{2}\left(\sum_{n=1}^{N} y_n - w_1 \sum_{n=1}^{N} x_n\right)}{\cancel{2}N} = w_0$$

$$1/N \sum_{n=1}^{N} y_n - w_1\left(1/N \sum_{n=1}^{N} x_n\right) = w_0$$

$$\frac{\partial(w_1)}{w_1} = \sum_{n=1}^{N} 2(y_n - w_0 - w_1 x_n) \cdot (0 - 0 - x_n)$$

$$0 = \sum_{n=1}^{N} -2x_n (y_n - w_0 - w_1 x_n)$$

$$0 = \sum_{n=1}^{N} -2x_n y_n + 2x_n w_0 + 2w_1 x_n^2$$

$$-2w_1 \sum x_n^2 = 2w_0 \sum x_n - 2\sum_{n=1}^{N} x_n y_n$$

$$w_1 = \frac{\cancel{2}(w_0 \sum x_n - \sum x_n y_n)}{-\cancel{2}\sum x_n^2} \Rightarrow w_1 = \frac{w_0 \sum x_n - w_1 \sum x_n}{\sum x_n^2}$$

b)

If $\bar{x} = 1/N \sum_{n=1}^{N} x_n$ and $\bar{y} = 1/N \sum_{n=1}^{N} y_n$

$$W_0^* = \left(1/N \sum_{n=1}^{N} y_n\right) - w_1 \left(1/N \sum_{n=1}^{N} x_n\right)$$

$$= \bar{y} - w_1 \bar{x}$$

$$W_1^* = \frac{\sum_{n=1}^{N} x_n y_n - N\left(1/N \sum_{n=1}^{N} x_n\right)\left(1/N \sum_{n=1}^{N} y_n\right)}{\sum_{n=1}^{N} x_n^2 - N\left(1/N \sum_{n=1}^{N} x_n\right)^2}$$

$$= \frac{x_n y_n - N(\bar{x})(\bar{y})}{x_n^2 - N(\bar{x})^2} - \frac{N(\bar{x})(\bar{y})}{x_n^2 - N(\bar{x})^2}$$

$$= \frac{\sum_{n=1}^{N} x_n y_n}{\sum_{n=1}^{N} x_n^2 - N(\bar{x})^2} - \frac{x_n \bar{y}}{x_n (x_n - \bar{x})}$$

$$= \frac{\sum_{n=1}^{N} x_n y_n}{\sum_{n=1}^{N} x_n^2 - N\left(1/N \sum_{n=1}^{N} x_n\right)\left(1/N \sum_{n=1}^{N} x_n\right)} - \frac{\bar{y}}{(x_n - \bar{x})}$$

$$= \frac{\sum x_n y_n}{\sum_{n=1}^{N} x_n^2 - x_n \cdot \bar{x}} - \frac{\bar{y}}{(x_n - \bar{x})}$$

$$W_1^* = \frac{\sum y_n}{\sum (x_n - \bar{x})} - \frac{\sum \bar{y}}{\sum (x_n - x)}$$

$$= \frac{\sum_{n=1}^{N} y_n - \bar{y}}{\sum_{n=1}^{N} (x_n - \bar{x})}$$

$$= \frac{\sum_{n=1}^{N} y_n - \bar{y}}{\sum_{n=1}^{N} (x_n - \bar{x})} \cdot \frac{(x_n - \bar{x})}{(x_n - \bar{x})}$$

$$= \frac{\sum_{n=1}^{N} (y_n - \bar{y})(x_n - \bar{x})}{\sum_{n=1}^{N} (x_n - \bar{x})^2}$$

c) W_1* represents the covariance of the variances of the input features "x" and "y". W_0* represents the difference of the means of x and y associated with the D^{train}. W_1* is used as a weight to influence the mean of x as adjusted for minimized prediction error.

Question 2:

a) The gradient vector of J with respect to x(0) is a first-order derivative while the Hessian matrix, which contains the second order derivatives, specifically the derivative with respect to x(0). In accordance to the Taylor expansion to the second order, the function is approximated as follows:

$$f(x) = f(x_0) + (\nabla f|_{x=x_0})^T \cdot (x - x_0) + \tfrac{1}{2} (x - x_0)^T \cdot H_f|_{x=x_0} \cdot (x - x_0)$$

$$J(w) \approx J(w(k)) + (\nabla J|_{w=w(k)})^T \cdot (w - w(k)) + \tfrac{1}{2} (w - w(k))^T \cdots$$
$$\cdot H_J|_{w=w(k)} \cdot (w - w(k))$$

b)

$$J(w(k+1)) \approx J(w(k)) + \left(\nabla J\big|_{w=w(k+1)}\right)^T \cdot (w(k) - w(k+1))$$
$$+ \tfrac{1}{2}(w(k) - w(k+1))^T \cdot H_J\big|_{w=w(k)} \cdot (w - w(k))$$

$$\approx J(w(k)) + \|\nabla J\|_{w=w(k)}\|_2^2 \cdot (\cancel{w(k)} - \cancel{w(k)} - a(k)) \cdot \nabla J\big|_{w=w(}$$
$$+ \tfrac{1}{2}\left((\cancel{w(k)} - \cancel{w(k)} - a(k) \cdot J\big|_{w=w(k)}\right)^T \cdot H_J\big|_{w=w(k)} \cdots$$
$$\cdots \left(\cancel{w(k)} - \cancel{w(k)} - a(k) \cdot \nabla J\big|_{w=w(k)}\right)$$

$$\approx J(w(k)) + \|\nabla J\big|_{w=w(k)}\|_2^2 \cdot a(k) + \ldots$$
$$+ \tfrac{1}{2}\left((\nabla J\big|_{w=w(k)})^T (H_J\big|_{w=w(k)}) (\nabla J\big|_{w=w(k)}) \cdot a^2(k)\right)$$

c)

$$J(w(k)) - \|\nabla J\|$$

$$J(w(k)) - \|\nabla J|_{w=w(k)}\|_2^2 \cdot 1 + \tfrac{1}{2}($$

$$J(w) \approx J(w(k)) - \|\nabla J|_{w=w(k)}\|_2^2 \cdot a(k) + \tfrac{1}{2}(A) \cdot a^2(k)$$

Let $A = \left(\left(\nabla J|_{w=w(k)}\right)^T H_J|_{w=w(k)} \left(\nabla J|_{w=w(k)}\right)\right)$

$$J'(w) \approx -\|\nabla J|_{w=w(k)}\|_2^2 \cdot 1 + \tfrac{1}{2}(A) \cdot a^2(k)$$

$$\approx -\|\nabla J|_{w=w(k)}\|_2^2 + \cancel{\tfrac{1}{2}}(A) \cdot \cancel{2}a(k)$$

$$\approx -\|\nabla J|_{w=w(k)}\|_2^2 + A \cdot a(k)$$

$$\|\nabla J|_{w=w(k)}\|_2^2 = A \cdot a(k)$$

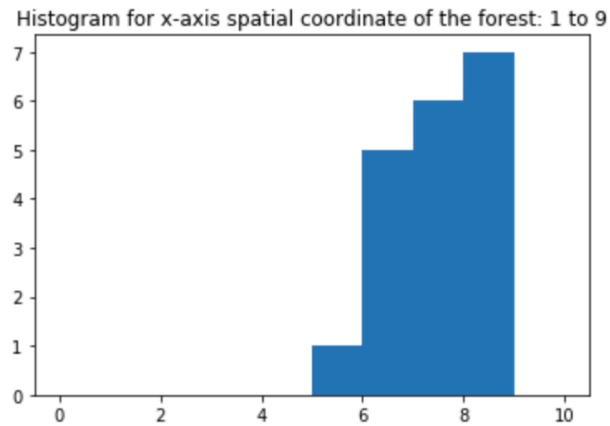
$$\frac{\|\nabla J|_{w=w(k)}\|_2^2}{A} = a(k)$$

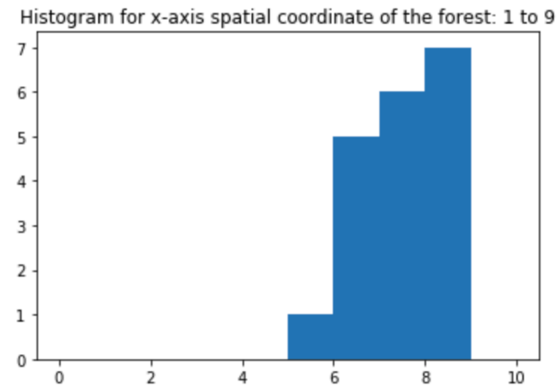
$$a(k) = \frac{\|\nabla J|_{w=w(k)}\|_2^2}{\left(\nabla J|_{w=w(k)}\right)^T \left(H_J|_{w=w(k)}\right) \left(\nabla J|_{w=w(k)}\right)}$$

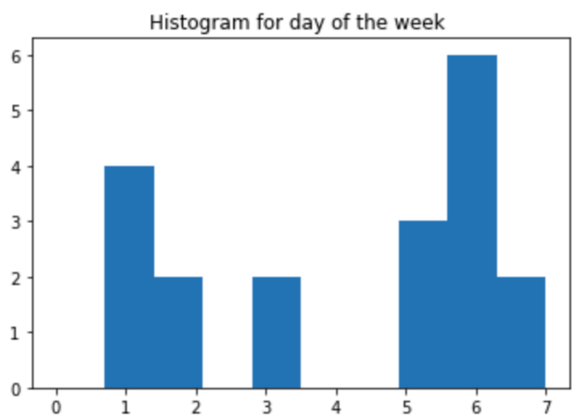
Question 3:

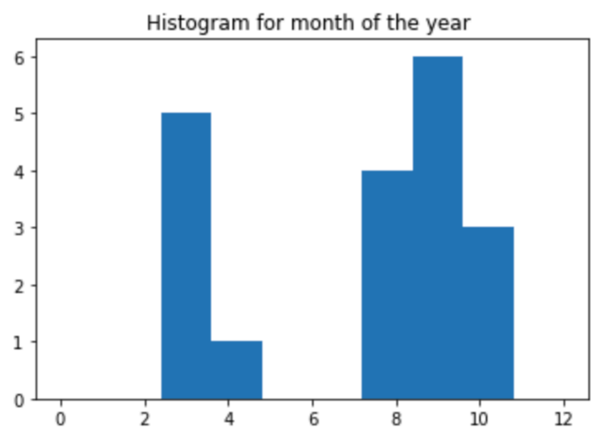
a) The following features are categorical: ‘X’, ‘Y’, ‘month’, and ‘day’. The following features are continuous: ‘FFMC’, ‘DMC’, ‘DC’, ‘ISI’, ‘temp’, ‘RH’, ‘wind’, and ‘rain’.

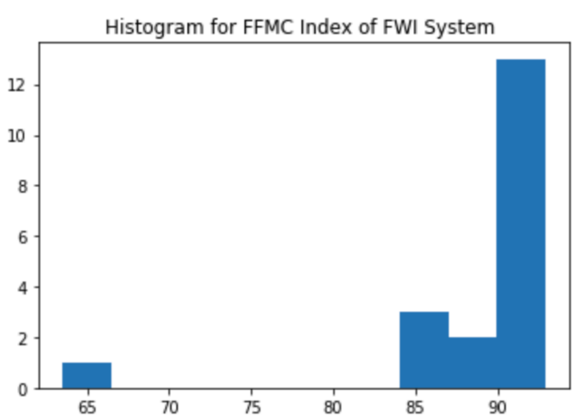
Histograms:

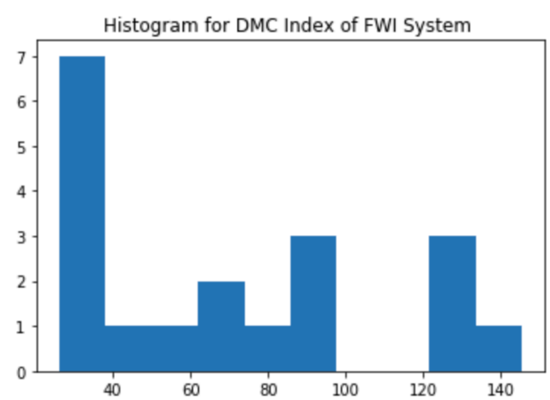

Histogram for x-axis spatial coordinate of the forest: 1 to 9

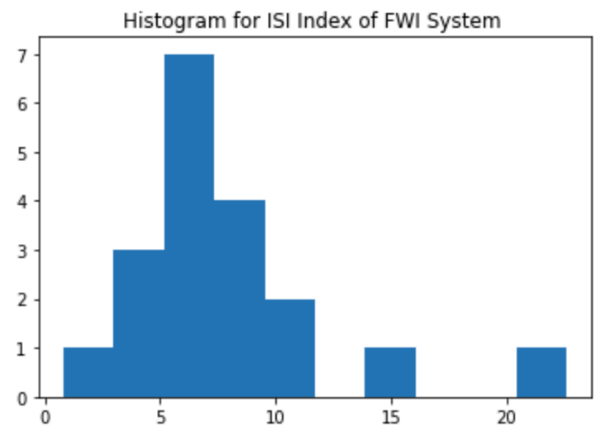

Histogram for x-axis spatial coordinate of the forest: 1 to 9

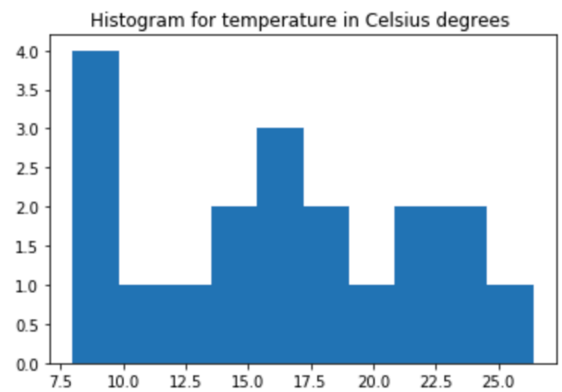

Histogram for day of the week

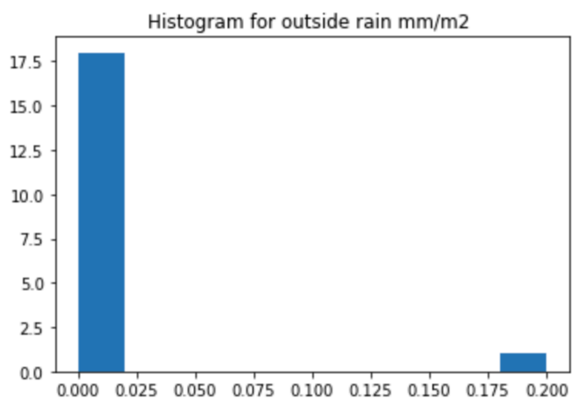

Histogram for month of the year

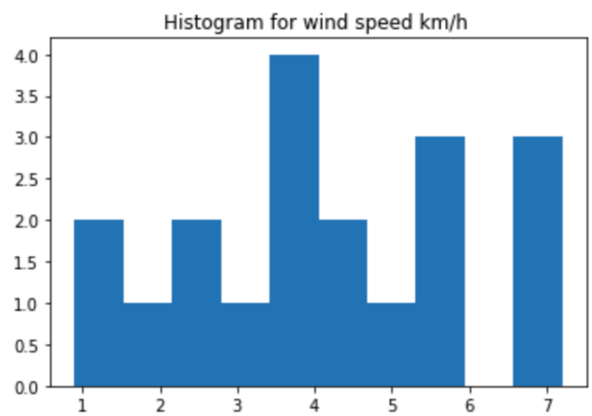

Histogram for FFMC Index of FWI System


Histogram for DMC Index of FWI System









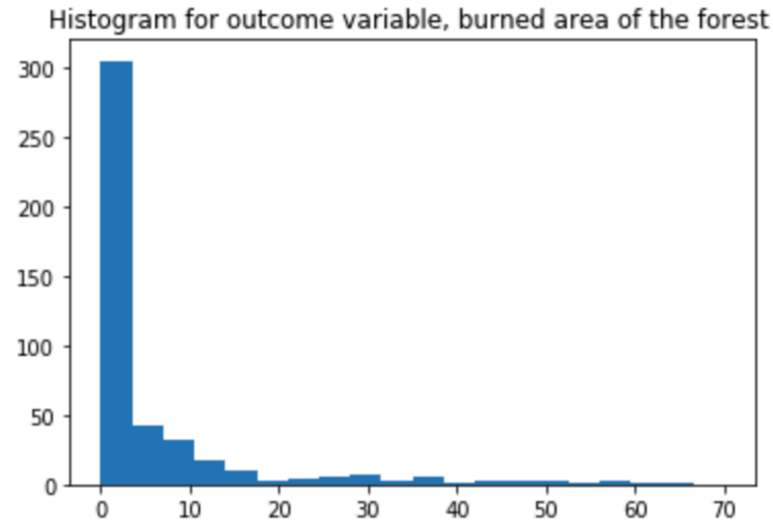
b-i)

b-ii)

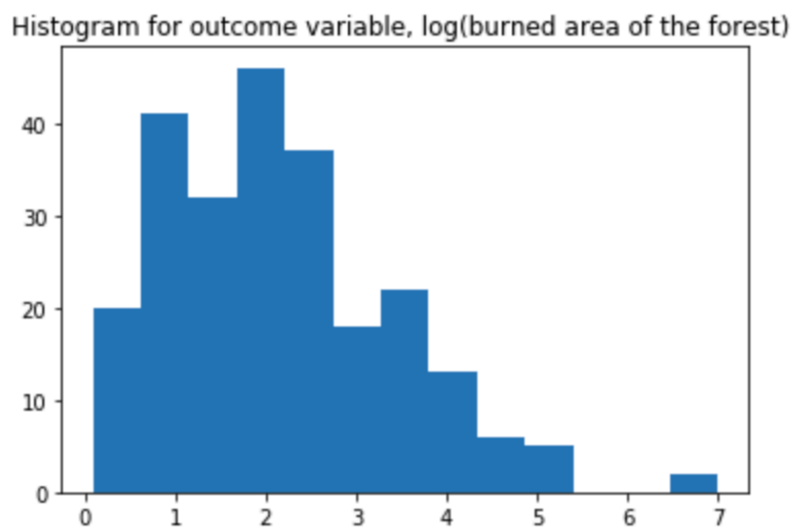
b-iii)

c)

Histogram of outcome variables:



Outcome values appear to be concentrated between 10-17 area units.
Histogram of logarithm of outcome variables:



Outcome values appear to be concentrated between 1-7 area units. Frequencies appeared to have decreased as they were distributed in the space.