10

1. (a) target variable: average shooting score in soccer games in the last year. that the player took.

Explain: must be record of the shooting score for players in soccer games last year. for a soccer team.

(b) ~~height, weight~~, training hours, shooting scores in the training process.

(c) Yes, a linear model would be reasonable

I assume the slope to be positive.

2. (a) $N = 10^{\gamma - 2M}$

$\log_{10} N = \log_{10} 10^{\gamma - 2M}$

$\log_{10} N = \gamma - 2M$

let $y = \log_{10} N$

$x = M$

$\beta_0 = \gamma$

$\beta_1 = -2$

$\longrightarrow y = \beta_0 + \beta_1 x$

(b) To compute the least-squares estimates of $\{\beta_0, \beta_1\}$ we should compute $\{\beta_0, \beta_1\}$ that minimize residual sum of squares (RSS):

$$RSS(\beta_0, \beta_1) \triangleq \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

To minimize $RSS(\beta_0, \beta_1)$ we find the β_0 and β_1 that zero the gradient.

$$\frac{\partial RSS(\beta_0, \beta_1)}{\partial \beta_0} = 0 \qquad \frac{\partial RSS(\beta_0, \beta_1)}{\partial \beta_1} = 0$$

$$0 = \frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \qquad -2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$0 = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) \qquad \frac{1}{n} \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$0 = \underbrace{\frac{1}{n} \sum_{i=1}^{n} y_i}_{\bar{y}} - \beta_0 - \beta_1 \underbrace{\frac{1}{n} \sum_{i=1}^{n} x_i}_{\bar{x}} \qquad \frac{1}{n} \sum_{i=1}^{n} x_i y_i - \bar{x}\beta_0 - \beta_1 \frac{1}{n} \sum_{i=1}^{n} x_i^2 = 0$$

$$\underbrace{\frac{1}{n} \sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}_{S_{xy}} - \beta_1 \underbrace{\left(\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2\right)}_{S_{xx}}$$

$$\Longleftrightarrow \beta_0 = \bar{y} - \beta_1 \bar{x} \qquad \Longleftrightarrow \beta_1 = \frac{S_{xy}}{S_{xx}}$$

(c) from (b) we estimated β_0, β_1

where $\beta_0 = \bar{y} - \beta_1 \bar{x}$

$$\beta_1 = \frac{S_{xy}}{S_{xx}} \quad \left(S_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} \right.$$

$$\left. S_{xx} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 \right)$$

$$\gamma = \beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\alpha = -\beta_1 = -\frac{S_{xy}}{S_{xx}}$$

4. (a) $RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \beta x_i)^2$

(b) we should compute β that minimize RSS. we find β that zero the gradient:

$$\frac{\partial RSS}{\partial \beta} = 0$$

$$\frac{\partial}{\partial \beta} \sum_{i=1}^{n} (y_i - \beta x_i)^2 = 0$$

$$\cancel{-2x_i \sum_{i=1}^{n} (y_i - \beta x_i) = 0}$$

$$-2\sum_{i=1}^{n} x_i (y_i - \beta x_i) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i (y_i - \beta x_i) = 0$$

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i y_i}_{S_{xy} + \bar{x}\bar{y}} = \beta \cdot \underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i^2}_{S_{xx} + \bar{x}^2}$$

$$\beta = \frac{S_{xy} + \bar{x}\bar{y}}{S_{xx} + \bar{x}^2}$$