HW 2

1.

a) The students' GPA will be the target.

b) The target is continuous

c) One possible variable will be their high school GPA, SAT or ACT score.

d) Yes a linear model is reasonable, if using a variable of high school GPA. I expect the slope to be close to $w = 1$, since the better a student do in high school, it is reasonable to expect they also perform well in university.

2.

a)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 2$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 6$$

b)

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

$$S_{yy} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = 37.2$$

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 8$$

c)

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

$$\beta_1 = \frac{S_{xy}}{S_{xx}} = \frac{8}{2} = 4$$

$$\beta_0 = \bar{y} - \beta\bar{x} = -2$$

$$\epsilon = MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 5.2$$

d)

$$\hat{y} = 4x - 2$$

$$\hat{y}(2.5) = 4 \cdot 2.5 - 2 = 7$$

3.

$$z(t) \approx z_0 e^{-\alpha t}$$

a)

$$\ln(z(t)) \approx \ln(z_0 e^{-\alpha t})$$

$$\approx \ln(z_0) + \ln(e^{-\alpha t})$$

$$\approx \ln(z_0) + (-\alpha t \cancel{\ln(e)}\ 1)$$

$$\underbrace{\ln(z(t))}_{y} \approx \underbrace{-\alpha}_{\beta_1} \underbrace{t}_{x} + \underbrace{\ln(z_0)}_{\beta_0}$$

b)

$$y = \ln(z(t))$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y = \frac{1}{n}\sum_{i=1}^{n} \ln(z(t))$$

$$\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t$$

$$S_{tt} = \frac{1}{n}\sum_{i=1}^{n}(t-\bar{t})^2$$

$$S_{yt} = \frac{1}{n}\sum_{i=1}^{n}(t-\bar{t})(y-\bar{y})$$

$$S_{yy} = \frac{1}{n}\sum_{i=1}^{n}(y-\bar{y})^2$$

$$-\alpha = \frac{S_{yt}}{S_{tt}} \qquad \alpha = -\frac{\frac{1}{n}\sum_{i=1}^{n}(t-\bar{t})\left(\ln(z(t)) - \frac{1}{n}\sum_{i=1}^{n}\ln(z(t))\right)}{\frac{1}{n}\sum_{i=1}^{n}(t-\bar{t})^2}$$

$$\ln(z_0) = \bar{y} - \alpha\bar{t}$$

$$z_0 = e^{\bar{y}-\alpha\bar{t}}$$

c)

```
def fit_model(t, z):
    tm = np.mean(t)
    y = np.log(z)
    ym = np.mean(y)

    stt = np.mean((t-tm)**2)
    syt = np.mean((t-tm)*(y-ym))

    alpha =-syt/ stt

    z_zero = ym-alpha*tm
    return alpha, z_zero
```

Another example (done in class)

$$\hat{y} = \frac{1}{ax+b} \qquad \frac{1}{\hat{y}} = ax+b$$

Code :

```
def  fit_model (x,y)
   xm =  np.mean(x)
    z  =  1/y
   zm =  np.mean(z)
   sxx  =  np.mean((xm - x)^2)
   sxz  =  np.mean((zm - z)^2)
   szz  =  np.mean((x-xm) * (z-zm))

   beta1  =  sxz/sxx
   beta0  =  zm - beta1 * xm
   return  a, b
```

This solution minimizes

$$L = \sum \left(\frac{1}{y_i} - ax_i - b \right)^2$$

Not the same as:

$$L = \sum \left(y_i - \frac{1}{ax_i + b} \right)^2$$

4.

a) $$RSS(\beta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_i x_i)^2$$

b) $$L = \frac{1}{n}\sum(y_i - \beta x_i)^2$$

$$\frac{dL}{d\beta} = \frac{2}{n}\sum(y_i - \beta x_i)\left(\overset{0}{\cancel{\frac{dy_i}{d\beta}}} - x_i\right)$$

$$\frac{dL}{d\beta} = \frac{2}{n}\sum(y_i - \beta x_i)(-x_i)$$

$$0 = \frac{2}{n}\sum(-y_i x_i + \beta x_i^2)$$

$$= \frac{2}{n}\sum(-y_i x_i) + \frac{2}{n}\beta\sum x_i^2$$

$$+\beta\sum x_i^2 = +\sum(y_i x_i)$$

$$\beta = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}$$