1.

a) Students' accumulative GPA can be a possible target variable.

b) It is continuous.

c) Students' GRE score. Students' ILETS or TOEFL score can also be a variable.

d) A linear model should be reasonable for the data. I expect the model as a increasing linear function. The student with higher GRE score should have higher GPA.

2.

a) x_bar = (0 + 1 + 2 + 3 + 4) / 5 = 2
y_bar = (0 + 2 + 3 + 8 + 17) / 5 = 6

b) s_x^2 = 2.5
s_y^2 = 46.5
s_xy = 8

c) The least squares parameters are beta_0 and beta_1.
beta_1 = s_xy / s_x^2 = 3.2
beta_0 = y_bar – beta_1 * x_bar = -0.4

d) When x = 2.5, y = 7.6

3.

a)
$\ln z(t) \approx \ln z_0 e^{-\alpha t}$
$\ln z(t) \approx -\alpha t + \ln z_0$
Now z_0 and a appeared linearly.

b) RSS = $\sum_{i=1}^{n}(\ln z(t)_i - \ln z(t_i))^2 = \sum_{i=1}^{n}(\ln z(t)_i + \alpha t_i - \ln z_0)^2$
By replacing $\ln z(t)_i$ with y_i, t_i with $-x_i$, $\ln z_0$ with β_0 and α with β_1, we got RSS = $\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$, which is exactly the same as the original model.
We know the solution to the original model are
$\beta_1 = \frac{S_{xy}}{S_{xx}},\ \beta_0 = \bar{y} - \beta_1 \bar{x}$
By replacing variables back, we got
$\alpha = -\frac{S_{t \ln z(t)}}{S_{tt}},\ \ln z_0 = \overline{\ln z(t)} + \alpha \bar{t},\ z_0 = \overline{z(t)} + e^{\alpha t}$

c)

```python
import numpy as np

def fit_linear(t, zt):
    x = -t
    y = zt

    xm = np.mean(x)
    ym = np.mean(y)
    syx = np.mean((y - ym) * (x - xm))
    sxx = np.mean((x - xm) ** 2)
    beta1 = syx/sxx
    beta0 = ym - beta1 * xm
    yhat = beta0 + beta1 * x
    RSS = np.sum((y - yhat) ** 2)

    alpha = beta1
    lnz0 = beta0
    z0 = np.exp(lnz0)
    return alpha, z0, RSS
```

4.

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

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

When $\frac{\partial RSS}{\partial \beta} = 0$, RSS has the minimum value.

Let $\frac{\partial RSS}{\partial \beta} = 0,\ \beta = \sum_{i=1}^{n} \frac{y_i}{x_i}$