4.2 The data in Table 4.6 are times to death, y_i, in weeks from diagnosis and $\log_{10}$(initial white blood cell count), x_i, for seventeen patients suffering from leukemia. (This is Example U from Cox and Snell 1981.)

Table 4.6 *Survival time, y_i, in weeks and $\log_{10}$(initial white blood cell count), x_i, for seventeen leukemia patients.*

y_i	65	156	100	134	16	108	121	4	39
x_i	3.36	2.88	3.63	3.41	3.78	4.02	4.00	4.23	3.73
y_i	143	56	26	22	1	1	5	65	
x_i	3.85	3.97	4.51	4.54	5.00	5.00	4.72	5.00	

(a) Plot y_i against x_i. Do the data show any trend?

(b) A possible specification for $\mathrm{E}(Y)$ is

$$\mathrm{E}(Y_i) = \exp(\beta_1 + \beta_2 x_i),$$

which will ensure that $\mathrm{E}(Y)$ is non-negative for all values of the parameters and all values of x. Which link function is appropriate in this case?

(c) The Exponential distribution is often used to describe survival times.

The probability distribution is $f(y;\theta) = \theta e^{-y\theta}$. This is a special case of the Gamma distribution with shape parameter $\phi = 1$ (see Exercise 3.12(a)). Show that $\mathrm{E}(Y) = 1/\theta$ and $\mathrm{var}(Y) = 1/\theta^2$.

(d) Fit a model with the equation for $\mathrm{E}(Y_i)$ given in (b) and the Exponential distribution using appropriate statistical software.

(e) For the model fitted in (d), compare the observed values y_i and fitted values $\widehat{y}_i = \exp(\widehat{\beta}_1 + \widehat{\beta}_2 x_i)$, and use the standardized residuals $r_i = (y_i - \widehat{y}_i)/\widehat{y}_i$ to investigate the adequacy of the model. (Note: $\widehat{y}_i$ is used as the denominator of r_i because it is an estimate of the standard deviation of Y_i—see (c) above.)