Exercise sheet 1 solutions

1. We have already seen the formulae for these answers, so we just plug in the numbers. The MLE is
$$\hat{q} = \frac{k}{n} = \frac{18}{200} = 0.09.$$
The (estimated) standard error is given by
$$se(\hat{q}) = \sqrt{\frac{\hat{q}(1-\hat{q})}{n}} = \sqrt{\frac{0.09 \times (1-0.09)}{200}} = 0.0202.$$

2. The overall likelihood is the joint probability density function
$$p(t \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda S}.$$
where $S = \sum_{i=1}^{n} t_i$.
Hence the log-likelihood is
$$\ell(\lambda; t) = n \log(\lambda) - \lambda S.$$
Differentiating and setting to zero gives
$$\frac{d\ell}{d\lambda} = \frac{n}{\lambda} - S = 0.$$
Hence the MLE is
$$\hat{\lambda} = \frac{n}{S}.$$
For the data in the question, $n = 6$ and $S = 55$, so we have
$$\hat{\lambda} = \frac{6}{55} = 0.109 \text{ days}^{-1}.$$
For the second part, we saw in the lectures that the likelihood function is
$$p(t \mid \lambda) = \lambda^m e^{-\lambda S}.$$
where m is the number of deaths we observed. The derivation of the MLE is as above but with m instead of n, leading to
$$\hat{\lambda} = \frac{m}{S}.$$
Here $m = 5$ and S has changed to 56, and so
$$\hat{\lambda} = \frac{5}{56} = 0.0893 \text{ days}^{-1}.$$

3. The likelihood contribution for each data-point is the normal $N(\mu_i, \sigma^2)$ pdf. The overall likelihood is the joint probability density function

$$p(y \mid \beta_0, \beta_1, \sigma) = \prod_{i=1}^{n} \phi(y_i \mid \mu_i, \sigma).$$

where $\phi(...)$ is the normal pdf.

The log-likelihood (similar to the normal non-regression example in the lectures, but with μ_i instead of μ) is

$$\begin{aligned}
\ell(\beta_0, \beta_1, \sigma; y) &= -n\log(\sqrt{2\pi}) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu_i)^2 \\
&= -n\log(\sqrt{2\pi}) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2
\end{aligned}$$

The first derivatives of the log-likelihood are

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta_0} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) \\
\frac{\partial \ell}{\partial \beta_1} &= \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i) \\
\frac{\partial \ell}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2
\end{aligned}$$

Note that setting $\frac{\partial \ell}{\partial \beta_0} = 0$ and $\frac{\partial \ell}{\partial \beta_1} = 0$ gives the same estimates for β_0 and β_1 as setting $\frac{\partial S}{\partial \beta_0} = 0$ and $\frac{\partial S}{\partial \beta_1} = 0$, where S is the sum of squares $S = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$. Hence the MLEs $\hat{\beta}_0$ and $\hat{\beta}_1$ are the same as the least squares estimates found in Statistical Modelling 1. However, here we are maximizing the log-likelihood, whereas we minimize S.

The details for these two parameters are as follows.

$$\frac{\partial \ell}{\partial \beta_0} = 0 \implies \sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_i = 0 \implies \bar{y} - \beta_0 - \beta_1 \bar{x} = 0$$

So $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.

$$\frac{\partial \ell}{\partial \beta_1} = 0 \implies \sum_{i=1}^{n} x_i y_i - n\beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0 \implies \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \beta_0 \bar{x} - \frac{\beta_1}{n}\sum_{i=1}^{n} x_i^2 = 0$$

Substituting $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ gives

$$\frac{1}{n} \sum_{i=1}^{n} x_i y_i - (\bar{y} - \hat{\beta}_1 \bar{x})\bar{x} - \frac{\hat{\beta}_1}{n} \sum_{i=1}^{n} x_i^2 = 0$$

$$\hat{\beta}_1 = \frac{\frac{1}{n} \sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2}$$

This value can then be substituted into $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.

Then these values can both be substituted into $\dfrac{\partial \ell}{\partial \sigma} = 0$ to give MLE

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$$

As with the non-regression normal example, the MLE $\hat{\sigma}^2$ has a factor $1/n$, whereas the unbiased estimate of σ^2 which you came across in Statistical Modelling 1 has a factor $1/(n-2)$ for this regression model with one covariate.