Problem 1.

(a) Firstly the likelihood function is formed by computing

$$L(\theta) = f(x_1, \dots, x_n | \theta)$$

$$= \prod_{i=1}^{n} f(x_i | \theta) \quad \text{where } \theta = (\mu, \sigma) \text{ in this case.}$$

The log-likelihood is $\ell(\theta) = \log L(\theta)$.

The maximum likelihood estimate is found by maximizing $\ell(\theta)$ and it satisfies

$$\ell(\hat{\theta}) \geq \ell(\theta) \quad \text{for all } \theta.$$

The maximization can be done using calculations or using numerical methods (eg. Newton Raphson).

(b)

We know that

$\hat{\mu} \pm 1.96\sqrt{I_{11}^{-1}}$ is an approximate

95% confidence interval for μ.

Likewise $\hat{\sigma} \pm 1.96\sqrt{I_{22}^{-1}}$.

In this case we get

μ: 54.73 ± 0.98

σ: 4.50 ± 0.71

(c) The normal distribution ignores the fact that the data cannot fall outside the range (30, 60). So, the model gives probability density over a greater range.

Ignoring this truncation will lead to under estimating σ and it could bias estimating μ (depending on the value of μ).

We can see that the $\hat{\sigma}$ is much smaller and the $\hat{\mu}$ is a little smaller

(d) AIC (Truncated Normal) $= 2\ell - 2p$
$= -1068$

AIC (Normal) $=$ $2\ell - 2p = -1096$

AIC prefers the truncated normal (so does BIC)