Q1

(a)

Given $X_1, X_2, ..., X_n \sim N(\theta, \sigma^2)$ — σ is known.

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $se^2 = \sigma^2/n$

Prior distribution for $\theta \sim N(a, b^2)$. We have,

$\text{f}(\theta) = (2\pi b^2)^{\frac{-1}{2}}.exp(\frac{-(\theta-a)^2}{2b^2}) \quad -- \mathbf{1}$
$f(\mathbf{x}|\theta) = (2\pi\sigma^2)^{\frac{-1}{2}}.\Pi_{i=1}^{n} exp(\frac{-(x_i-\theta)^2}{2\sigma^2}) \quad -- \mathbf{2}$
$f(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta).f(\theta)$
$Using\ 1\ and\ 2;$
$f(\theta|\mathbf{x}) = (2\pi b^2)^{\frac{-1}{2}}.exp(\frac{-(\theta-a)^2}{2b^2}).(2\pi\sigma^2)^{\frac{-1}{2}}.\Pi_{i=1}^{n} exp(\frac{-(x_i-\theta)^2}{2\sigma^2})$
$= exp(\frac{-1}{2}\{\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{\sigma^2} + \frac{(\theta-a)^2}{b^2}\})$
$= exp(\frac{-1}{2}\{\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i^2 + \theta^2 - 2x_i\theta) + \frac{(\theta-a)^2}{b^2}\})$
$Ignoring\ constants;$
$= exp(-\frac{\theta^2 n}{2\sigma^2} + \frac{2\theta\sum_{i=1}^{n} x_i}{\sigma^2} - \frac{\theta^2}{2b^2} - \frac{a^2}{2b^2} + \frac{\theta a}{b^2})$
$= exp(\theta^2(-\frac{n}{2\sigma^2} - \frac{1}{2b^2}) + \theta(\frac{n\overline{X}}{\sigma^2} + \frac{a}{b^2}) + constant) \quad -- \mathbf{3}$
$For\ a\ Normal\ distribution\ with\ response\ y\ with\ mean\ x\ and\ variance\ y^2\ we\ have$
$g(r) = (2\pi y^2)^{\frac{-1}{2}} exp\{(r-x)^2/2y^2\}$
$\propto exp\{\frac{-1}{2}r^2 y^{-1} + rx/y + constant\} \quad -- \mathbf{4}$
$Comparing\ equations\ \mathbf{3}\ and\ \mathbf{4}$
$x = y^2(\frac{a}{b^2} + \frac{n\overline{X}}{\sigma^2}) \quad -- \mathbf{5};$
$y^2 = (\frac{1}{b^2} + \frac{n}{\sigma^2})^{-1} \quad -- \mathbf{6}$
$Solving\ for\ y$
$y^2 = (\frac{1}{b^2} + \frac{1}{se})^{-1}$
$y^2 = \frac{b^2.se^2}{b^2+se^2} \quad -- \mathbf{7}$
$Putting\ \mathbf{7}\ in\ \mathbf{5};$
$x = \frac{b^2.se^2}{b^2+se^2}.\frac{b^2.\overline{X}+a.se^2}{b^2.se^2}$
$Thus,\ we\ have:\ x = \frac{b^2.\overline{X}+a.se^2}{b^2+se^2};\ y^2 = \frac{b^2.se^2}{b^2+se^2}$
Hence Proved!

(b)

Finding an interval $C = (c, d)$ such that $P(\theta \in C|\mathbf{x}) = (1-\alpha)$.

Choose c and d such that: $P(\theta < c|\mathbf{x}) = 0.025$ and $P(\theta > d|\mathbf{x}) = 0.025$

$$\begin{aligned} P(d < \theta < c|\mathbf{x}) &= P(\frac{(d-x)}{y} < \frac{(\theta-x)}{y} < \frac{(c-x)}{y}|\mathbf{x}) \\ &= P(\frac{(d-x)}{y} < Z < \frac{(c-x)}{y}) = (1-\alpha) \quad --\mathbf{I} \\ &\quad From\ definition\ of\ (1-\alpha)\ C.I; \\ &P(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}) = (1-\alpha) \quad --\mathbf{II} \\ &Comparing \quad --\mathbf{I}\ and \quad --\mathbf{II} \\ &c = x + y.z_{\frac{\alpha}{2}}; \qquad d = x - y.z_{\frac{\alpha}{2}} \\ &Posterior\ interval = (x - y.z_{\frac{\alpha}{2}}, x + y.z_{\frac{\alpha}{2}}) \end{aligned}$$

Since $x \to \overline{X}\ and\ y \to se\ as\ n\ \to \infty$

$Posterior\ interval = (\overline{X} \pm z_{\frac{\alpha}{2}}.se)$

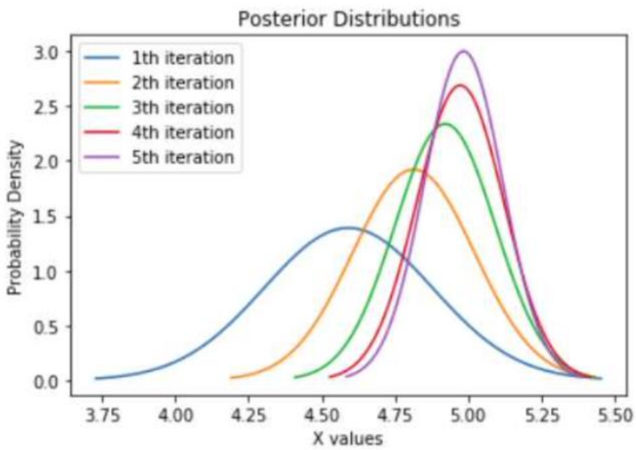
This is the frequentist confidence interval.

Q2

(a)

```
a6_q3('q2_sigma3.dat', 9)
```

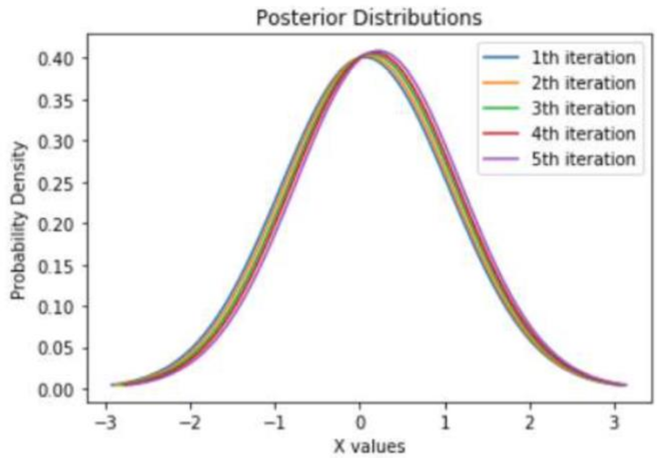
```
1th iteration: Mean = 4.590762414332327 Variance = 0.08256880733944953
2th iteration: Mean = 4.813523613446215 Variance = 0.0430622009569378
3th iteration: Mean = 4.921256878168492 Variance = 0.02912621359223301
4th iteration: Mean = 4.97283741207765 Variance = 0.022004889975550123
5th iteration: Mean = 4.983966097849453 Variance = 0.01768172888015717
```



(b)

```
a6_q3('q2_sigma100.dat', 10000)
```

```
1th iteration: Mean = 0.05871624147982311 Variance = 0.9900990099009901
2th iteration: Mean = 0.09500866961681816 Variance = 0.9803921568627452
3th iteration: Mean = 0.13822626152242073 Variance = 0.970873786407767
4th iteration: Mean = 0.17121883350740297 Variance = 0.9615384615384617
5th iteration: Mean = 0.2189182449674514 Variance = 0.9523809523809524
```



(c)

- When data has low variance, the posterior tends to converge i.e, move away from the prior.
- However, with high variance posterior remains close to the prior.

Q3

(a)

First we define the fitted equation to be an equation:

$$\hat{Y} = \beta_0 + \beta_1 X$$

Now, for each observed response Y_i, with a corresponding predictor variable X_i, so we would like to minimize the sum of the squared distances of each observed response to its fitted value.

$$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$$

Thus, we set the partial derivatives of $SSE(\beta_0, \beta_1)$ with respect β_0 and β_1 equal to zero

$$\begin{aligned}\frac{dSSE}{d\beta_0} &= \sum_{i=1}^{n} 2(-1)(Y_i - \beta_0 - \beta_1 X_i) = 0\\ &\Rightarrow \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i) = 0\end{aligned}$$

$$\begin{aligned}\frac{dSSE}{d\beta_1} &= \sum_{i=1}^{n} 2(-X_i)(Y_i - \beta_0 - \beta_1 X_i) = 0\\ &\Rightarrow \sum_{i=1}^{n} X_i(Y_i - \beta_0 - \beta_1 X_i) = 0\end{aligned}$$

The we could get 2 normal equations:

$$\beta_0 n + \beta_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i$$

$$\beta_0 \sum_{i=1}^{n} X_i + \beta_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i$$

For the first normal equation, we could get

$$\beta_0 = \frac{\sum_{i=1}^{n} Y_i - \beta_1 \sum_{i=1}^{n} X_i}{n}$$

Substitute into the second normal equation, yields,

$$\begin{aligned}
\frac{\sum_{i=1}^{n} Y_i - \beta_1 \sum_{i=1}^{n} X_i}{n} \sum_{i=1}^{n} X_i + \beta_1 \sum_{i=1}^{n} X_i^2 &= \sum_{i=1}^{n} X_i Y_i\\
\beta_1(\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}) &= \sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}\\
\beta_1(\sum_{i=1}^{n} X_i^2 - 2\frac{(\sum_{i=1}^{n} X_i)^2}{n} + \frac{(\sum_{i=1}^{n} X_i)^2}{n}) &= \sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n} - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n} + \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}\\
\beta_1(\sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n} X_i \frac{\sum_{i=1}^{n} X_i}{n} + \sum_{i=1}^{n}(\frac{\sum_{i=1}^{n} X_i}{n})^2) &= \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \bar{Y} - \sum_{i=1}^{n} Y_i \bar{X} + \sum_{i=1}^{n} \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n^2}\\
\beta_1 \sum_{i=1}^{n}(X_i^2 - 2X_i \frac{\sum_{i=1}^{n} X_i}{n} + (\frac{\sum_{i=1}^{n} X_i}{n})^2) &= \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \bar{Y} - \sum_{i=1}^{n} Y_i \bar{X} + \sum_{i=1}^{n} \bar{X}\bar{Y}\\
\beta_1 \sum_{i=1}^{n}(X_i^2 - \bar{X})^2 &= \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})\\
\Rightarrow \hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}
\end{aligned}$$

Thus we could have

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

(b)

First, we rewrite $\hat{\beta}_1$ as

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})Y_i}{S_{xx}} = \sum_{i=1}^n \frac{X_i - \bar{X})Y_i}{S_{xx}} = \sum_{i=1}^n c_i Y_i$$

and we could have $\sum_{i=1}^n c_i = \sum_i \frac{X_i - \bar{X}}{S_{xx}} = \frac{n\bar{X} - n\bar{X}}{S_{xx}} = 0$. Also, $E[\epsilon_i] = 0$. Then, we have

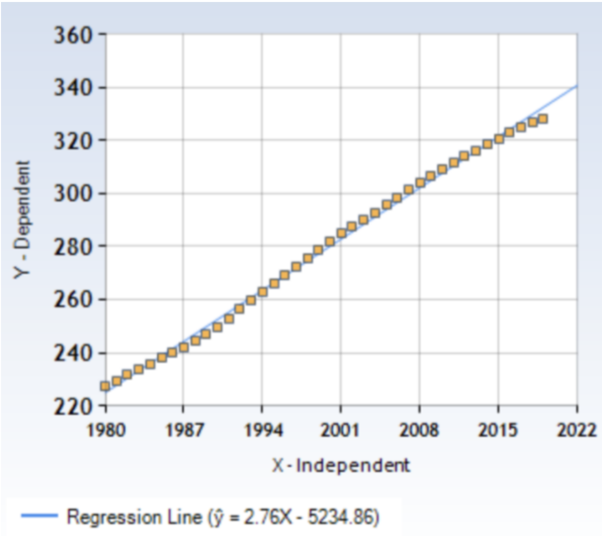
$$\begin{aligned}
E[\hat{\beta}_1] &= \sum_{i=1}^n c_i E[Y_i] \\
&= \sum_{i=1}^n c_i E[\beta_0 + \beta_1 X_i + \epsilon_i] \\
&= \beta_0 \sum_{i=1}^n c_i + \beta_1 \sum_{i=1}^n c_i X_i + \sum_{i=1}^n c_i E[\epsilon_i] \\
&= \beta_1 \sum_{i=1}^n \frac{(X_i - \bar{X})X_i}{\sum_{i=1}^n (X_i - \bar{X})^2} \\
&= \beta_1 \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \\
&= \beta_1
\end{aligned}$$

$$\begin{aligned}
E[\hat{\beta}_0] &= E[\bar{Y} - \hat{\beta}_1 \bar{X}] \\
&= E[\frac{\sum_{i=1}^n Y_i}{n} - \frac{\sum_{i=1}^n \hat{\beta}_1 X_i}{n}] \\
&= \frac{\sum_{i=1}^n E[\beta_0 + \beta_1 X_i - \hat{\beta}_1 X_i]}{n} \\
&= \frac{\sum_{i=1}^n (\beta_0 + \beta_1 X_i - \beta_1 X_i)}{n} \\
&= \beta_0
\end{aligned}$$

Q4

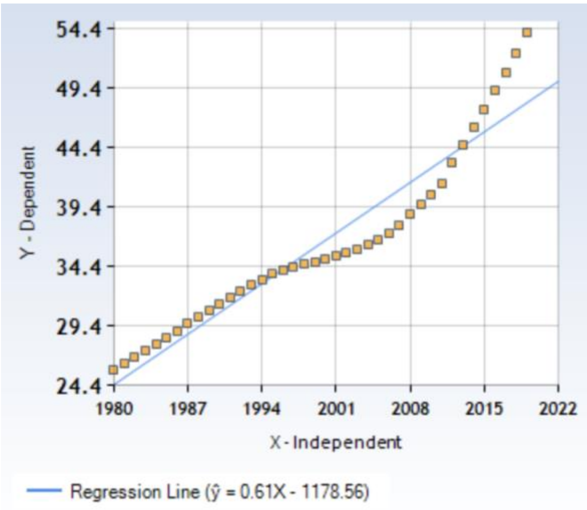
(a)

Total population V.s. year



SSE = 123.850183

65+ population V.s. year



SSE = 176.038626

Total population is more suitable for linear regression.

(b)

Using data from 1980-2018,

$$\hat{y} = 0.58376x - 1131.0736$$

SSE = 137.6942379, MSE = 3.5306214

When $x = 2060$, $\hat{y} = 71.4787$, when $x = 2019$, $|\hat{y} - y| = 6.5$

Using data from 2008-2018,

$$\hat{y} = 1.40254x - 2778.38941$$

SSE = 2.008894125, MSE = 0.18262674

When $x = 2060$, $\hat{y} = 110.8394$, when $x = 2019$, $|\hat{y} - y| = 0.722863$

You can use either MSE or prediction error in 2019 to show that using data from 2008-2018 is better. The media is right regarding the prediction in 2060. (Or underestimated it)

(c)

First way,

Ratio V.s. year, $\hat{y} = 0.00341x - 6.71902$.

When $x = 2019$, $\hat{y} = 0.1624$, $|\hat{y} - y|$ =0.0022462.

Second way,

$$ratio = \frac{\hat{y}(65 \uparrow \ population)}{\hat{y}(total\, population)} = \frac{53.33885}{329.6889} = 0.1617854$$

The error is 0.0028608.

The first way is more accurate. We already know total population is almost linear with year. The difference between these two methods is that for the first one we assume the ration is linear with year while the second assumption is 65+ population is linear with year. The first way is more accurate because the first assumption is more accurate, which means the ratio vs year is more linear. It will be more obvious if you use more data before 2008.

Q5

(a)

Admit chance = 0.00173741 * GRE + 0.00291958 * TOEFL + 0.00571666 * Rating - 0.00330517 * SOP + 0.02235313 * LOR + 0.11893945 * GPA + 0.02452511 * Research

SSE = 0.18431744

(b)

Admit chance = 0.01339112*TOEFL + 0.0145516*SOP + 0.04307257*LOR

SSE = 0.52099975

(c)

Admit chance = 0.00305067*GRE + 0.15994862*GPA

SSE = 0.2183461

(d)

It's not necessarily the case that the more attributes you have, more accurate your MLR model will be. (Or GPA is most significant feature.) (It actually depends on the 'quality' of the feature. The 'quality' evaluates how significant the relationship is between the dependable variable y and the feature. To know more about this, try to search for 'significance test' by yourself.)

Q6

(a)

$H \equiv RV$ for the soil type.

The two hypotheses are: $\boldsymbol{H_0}: H = 0$ and $\boldsymbol{H_1}: H = 1$ with $P(H = 0) = p$ and $P(H = 1) = (1 - p)$

Observations of water concentration metric $\boldsymbol{w} = \{w_1, \dots w_n\}$

$f_W(w|H = 0) = N(w; -\mu, \sigma^2)$ and $f_W(w|H = 1) = N(w; \mu, \sigma^2)$

Also $w_i s$ are conditionally independent of each other given the hypothesis/soil type.

$$P(H = 0|\boldsymbol{w}) = \frac{P(\boldsymbol{w}|H = 0)P(H = 0)}{P(\boldsymbol{w})} \qquad \textit{By Bayes theorem}$$

$$\Rightarrow \ P(H = 0|\boldsymbol{w}) = \frac{P(H=0)}{P(\boldsymbol{w})} \prod_{i=1}^{n} f_W(w_i|H = 0) \quad \because (w_i|H = h) \perp (w_i|H = h)$$

$$\Rightarrow \ P(H = 0|\boldsymbol{w}) = c.p.\exp\left(-\frac{\sum_i (w_i + \mu)^2}{2\sigma^2}\right)$$

We choose $H_0 (C = 0)$ if $P(H = 0|\boldsymbol{w}) \geq P(H = 1|\boldsymbol{w})$, i.e.

$$c.p.\exp\left(-\frac{\sum_i (w_i + \mu)^2}{2\sigma^2}\right) \geq c.(1 - p).\exp\left(-\frac{\sum_i (w_i - \mu)^2}{2\sigma^2}\right)$$

$$\Rightarrow \exp\left(-\frac{\sum_i (w_i + \mu)^2 - \sum_i (w_i - \mu)^2}{2\sigma^2}\right) \geq \frac{(1 - p)}{p}$$

$$\Rightarrow \exp\left(-\frac{2\mu \sum_i w_i}{\sigma^2}\right) \geq \frac{(1 - p)}{p}$$

$$\left(\sum_i w_i\right) \leq \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1 - p}\right)$$

(b)

For $P(H_0) = 0.1$, the hypothesis selected are: 0 1 0 0 1 0 1 1 0 1

For $P(H_0) = 0.3$, the hypothesis selected are: 0 1 0 0 1 0 1 1 0 1

For $P(H_0) = 0.5$, the hypothesis selected are: 0 1 0 0 1 0 1 1 0 1

For $P(H_0) = 0.8$, the hypothesis selected are: 0 1 0 0 1 0 1 1 0 1

(c)

We choose H_0 i.e. $C = 0$ iff

$$\left(\sum_i w_i\right) \leq \frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right)$$

We choose H_1 iff

$$\left(\sum_i w_i\right) > \frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right)$$

$$P(C=0|H=1) = P\left(\left(\sum_i w_i\right) \leq \frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right) | \, (H=1)\right)$$

$\because w_i|(H=0) \sim N(-\mu, \sigma^2)$

$$\Rightarrow \left(\sum_i w_i\right)|(H=0) \sim N(-n\mu, n\sigma^2)$$

$$\Rightarrow \left(\sum_i w_i\right)|(H=1) \sim N(n\mu, n\sigma^2)$$

$$\Rightarrow P(C=0|H=1) = \Phi\left(\frac{\frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right) - n\mu}{\sqrt{n\sigma^2}}\right) \quad \because if\ X \sim N(\mu, \sigma^2) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0,1)$$

Similarly,

$$P(C=1|H=0) = P\left(\left(\sum_i w_i\right) > \frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right) | \, (H=0)\right)$$

$$\Rightarrow P(C=1|H=0) = 1 - \Phi\left(\frac{\frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right) + n\mu}{\sqrt{n\sigma^2}}\right)$$

$$\therefore AEP = (1-p).\Phi\left(\frac{\frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right) - n\mu}{\sqrt{n\sigma^2}}\right) + p.\left(1 - \Phi\left(\frac{\frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right) + n\mu}{\sqrt{n\sigma^2}}\right)\right)$$