1. a)

$$\begin{aligned}
\mathrm{E}[\hat{\beta}|X] &= \mathrm{E}[(X'X)^{-1}X'Y|X] \\
&= \mathrm{E}[(X'X)^{-1}X'(X\beta + e)|X] \\
&= \mathrm{E}[(X'X)^{-1}X'X\beta + (X'X)^{-1}X'e|X] \\
&= \mathrm{E}[\beta|X] + (X'X)^{-1}X'\mathrm{E}[e|X] \\
&= \beta + (X'X)^{-1}X'b(X)
\end{aligned}$$

Thus, $\hat{\beta}$ is only unbiased when $(X'X)^{-1}X'b(X) = 0$. In other words, $\hat{\beta}$ is biased if $b(X)$ is in the span of $(X'X)^{-1}X'$.

b)

$$\begin{aligned}
\mathrm{Var}[\hat{\beta}|X] &= \mathrm{Var}[(X'X)^{-1}X'Y|X] \\
&= \mathrm{Var}[(X'X)^{-1}X'(X\beta + e)|X] \\
&= \mathrm{Var}[(X'X)^{-1}X'X\beta + (X'X)^{-1}X'e|X] \\
&= \mathrm{Var}[\beta + (X'X)^{-1}X'e|X] \\
&= \mathrm{Var}[(X'X)^{-1}X'e|X] \\
&= (X'X)^{-1}X'\mathrm{Var}[e|X]((X'X)^{-1}X')' \\
&= (X'X)^{-1}X'\sigma^2 I X(X'X)^{-1} \\
&= (X'X)^{-1}X'X(X'X)^{-1}\sigma^2 \\
&= (X'X)^{-1}\sigma^2
\end{aligned}$$

Thus, though $\hat{\beta}$ may be biased, it has no higher variance with assumption A2$'$ than with the usual assumption A2.

2. i. Given that the variable $wt_i = \beta_0 + e_i$, where each e_i is normally distributed around 0 and independent of all other e_j such that $j \neq i$, and independent of the variable height, the probability that a model that included height would reduce the residual sum of squares as much as it has or more is 17.31%.

   ii. Given that the variable $wt_i = \beta_0 + \beta_1 ht_i + e_i$, where each e_i is normally distributed around 0 and independent of all other e_j such that $j \neq i$, the probability that $|\hat{\beta}_0| \geq 36.8759$ given a null hypothesis that $\beta_0 = 0$ is 58.3%.

   iii. Given that the variable $wt_i = \beta_0 + \beta_1 ht_i + e_i$, where each e_i is normally distributed around 0 and independent of all other e_j such that $j \neq i$, the probability that $|\hat{\beta}_1| \geq .5821$ given a null hypothesis that $\beta_1 = 0$ is 17.3%.

   iv. You can not compute the probability of the fitted model being the right one, only the probability of seeing the given data and estimates on it for various assumptions of the underlying process that generated it.

3. All the individuals formulas for the desired quantities are:

$$SSreg = \frac{R^2 RSS}{1 - R^2} = \frac{.2084 \times 351191.4}{1 - .2084} = 92456.14$$

$$RSS = (RSE \times df)^2 = (2.993 \times 198)^2 = 351191.4$$

$$MSreg = \frac{SSreg}{regDF} = \frac{92456.14}{1} = 92456.14$$

$$\hat{\sigma}^2 = \frac{RSS}{resDF} = \frac{351191.41}{198} = 1773.694$$
$$F = \frac{MSreg}{\hat{\sigma}^2} = \frac{92456.14}{1773.694} = 1773.694$$

Which fit in the table as such:

Regression	df	ss	ms	F
Regression	regDF = 1	SSreg = 92456.14	MSreg = 92456.14	F = 52.126
Residual	resDf = 198	RSS = 351191.4	$\hat{\sigma}^2$ = 1773.694	

4. a) These are the resulting confidence intervals:

Confidence Level	β_0 LB	β_0 UB	β_1 LB	β_1 UP
66%	(3.802,	3.828)	(-.062,	-.059)
90%	(3.792,	3.837)	(-.063,	-.058)
95%	(3.788,	3.841)	(-.064,	-.057)

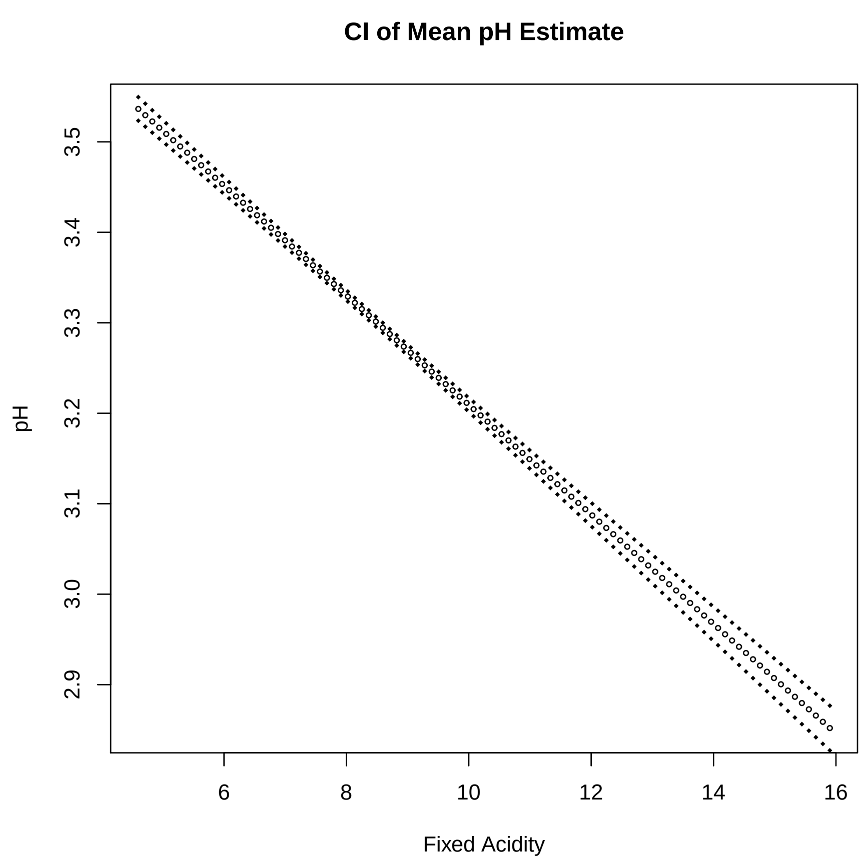
b) We get the variance estimate:

$$1.538 \times 10^{-5} = \sigma^2 \begin{bmatrix} 1 & 10 \end{bmatrix} \left(X'X\right)^{-1} \begin{bmatrix} 1 \\ 10 \end{bmatrix}$$

c) This is what we end up with, for a 95% CI:

Type	LB	Fit	UB
Mean	3.005	3.022	3.038
Prediction	2.800	3.022	3.245

d) The confidence interval is smallest near the mean of fixed acidity (8.31 in the data set). This stands to reason, since the confidence intervals are based on conditioned slices of a multivariate normal distribution; these slices will be steepest towards the center of the distribution (around the means of all variables), and flattest in the tails. In other words, its harder to screw up a prediction where you have a lot of data than where you don't have much data.



Here is my R code:

```
wine<-read.csv(file="hw1/winequalityRed.csv",head=TRUE,sep=";")
lm<-lm(pH ~ fixed.acidity, data=wine)

# Generate confidence intervals
confint(lm,level=.66)
confint(lm,level=.9)
confint(lm,level=.95)

# Generate variance estimate
c<-c(1,10)
n<-nrow(wine)
cov<-summary(lm)$cov.unscaled
sig2 <- summary(lm)$sigma^2
varpred<- sig2 * t(c) %*% cov %*% c
varpred

# Make prediction
predict(lm,newdata=data.frame(fixed.acidity=13.1), interval="confidence")
predict(lm,newdata=data.frame(fixed.acidity=13.1), interval="prediction")

# Make and plot 100 predictions
points<-data.frame(fixed.acidity =seq(min(wine$fixed.acidity),max(wine$fixed.acidity
predictions<-predict(lm,newdata=points, interval="confidence")
pdf('hw2_plot.pdf')
plot(points$fixed.acidity,predictions[,1],cex=.5,main="CI of Mean pH Estimate", xlab
points(points$fixed.acidity,predictions[,2],cex=.5, pch=18)
points(points$fixed.acidity,predictions[,3],cex=.5, pch=18)
dev.off()
```