Given $Y_i \overset{iid}{\sim} N(\mu_Y, \sigma_Y^2)$

1.a)

$$E[\bar{Y}] = E\left[\frac{1}{n}\sum Y_i\right] = \frac{1}{n}\sum E[Y_i] = \frac{1}{n}\cdot n\mu_Y = \mu_Y$$

$$Var[\bar{Y}] = Var\left(\frac{1}{n}\sum Y_i\right) = \frac{1}{n^2}\sum Var(Y_i) = \frac{1}{n^2}\cdot n\sigma_Y^2 = \frac{\sigma_Y^2}{n}$$

$$\frac{\bar{Y}-\mu_Y}{\sqrt{Var(\bar{Y})}} = \frac{\bar{Y}-\mu_Y}{\sqrt{\frac{\sigma_Y^2}{n}}} = \frac{\bar{Y}-\mu_Y}{\frac{\sigma_Y}{\sqrt{n}}} \sim N(0,1)$$

1.b)

$$E[\bar{Y}_1 - \bar{Y}_2] = E[\bar{Y}_1] - E[\bar{Y}_2] = \mu_1 - \mu_2$$

$$Var(\bar{Y}_1 - \bar{Y}_2) = Var(\bar{Y}_1) + Var(\bar{Y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

2.a) $f(y_i \mid \mu_Y, \sigma_Y^2) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{(y_i - \mu_Y)^2}{2\sigma_Y^2}\right)$

$$f(\mathbf{Y} \mid \mu_Y, \sigma_Y^2) = \prod_{i=1}^{n} f(y_i \mid \mu_Y, \sigma_Y^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{(y_i - \mu_Y)^2}{2\sigma_Y^2}\right)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_Y^2}}\right)^n \exp\left(-\frac{1}{2\sigma_Y^2}\sum_{i=1}^{n}(y_i - \mu_Y)^2\right)$$

$$= L(\mu_Y, \sigma_Y^2)$$

2.b)

$$\log L(\mu_Y, \sigma_Y^2 \mid \mathbf{Y}) = \log\left(\frac{1}{\sqrt{2\pi\sigma_Y^2}}\right) + \log\left(\exp\left(-\frac{1}{2\sigma_Y^2}\sum_{i=1}^{n}(y_i - \mu_Y)^2\right)\right)$$

$$= n\log\left(\frac{1}{\sqrt{2\pi\sigma_Y^2}}\right) - \frac{1}{2\sigma_Y^2}\sum_{i=1}^{n}(y_i - \mu_Y)^2$$

$$= -\frac{n}{2}\log(2\pi\sigma_Y^2) - \frac{1}{2\sigma_Y^2}\sum_{i=1}^{n}(y_i - \mu_Y)^2$$

2.c)

find $\hat{\mu}_{MLE}$:

$$\frac{d}{d\mu_Y}\log L(\mu_Y,\sigma_Y^2|Y) = \frac{d}{d\mu_Y}\left(\cancel{-\frac{n}{2}\log(2\pi\sigma_Y^2)}^{\,0} - \frac{1}{2\sigma_Y^2}\sum(y_i-\mu_Y)^2\right)$$

$$\left(\frac{-1}{2\sigma_Y^2}\cdot\left(-2\sum(y_i-\mu_Y)\right)\right) = \frac{1}{\sigma_Y^2}\sum(y_i-\mu_Y) = 0$$

$$0 = \sum(y_i-\mu_Y) \rightarrow \sum y_i = n\mu_Y \rightarrow \frac{1}{n}\sum y_i = \mu_Y$$

$$\hat{\mu}_{MLE} = \frac{1}{n}\sum y_i = \bar{Y}$$

find $\hat{\sigma}^2_{MLE}$:

$$\frac{d}{d\sigma_Y^2}\log L(\mu_Y,\sigma_Y^2|Y) = \frac{d}{d\sigma_Y^2}\left(-\frac{n}{2}\log(2\pi\sigma_Y^2) - \frac{1}{2\sigma_Y^2}\sum(y_i-\mu_Y)^2\right)$$

$$= \left(-\frac{n}{2}\cdot\frac{1}{\sigma_Y^2}\right) + \frac{1}{2(\sigma_Y^2)^2}\sum(y_i-\mu_Y)^2$$

$$\frac{-n}{2\sigma_Y^2} + \frac{1}{2(\sigma_Y^2)^2}\sum(y_i - \mu_Y)^2 = 0$$

$$\frac{1}{2(\sigma_Y^2)^2}\sum(y_i - \mu_Y)^2 = \frac{n}{2\sigma_Y^2}$$

$$\frac{1}{\sigma_Y^2}\sum(y_i - \mu_Y)^2 = n$$

$$\frac{1}{n}\sum(y_i - \mu_Y)^2 = \sigma_Y^2$$

$$\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum(y_i - \bar{Y})^2$$

1.23a)

No, $-0.027 = e$ → gotten from HW2-Problem 1.19 spreadsheet, cell P122

1.23b)

$$\sigma^2 = \frac{44.88}{120-2} = 0.38$$

→ Got from cell O122

$$\sigma = \sqrt{0.38} = 0.62$$

units are in GPA

1.28a)

$$\hat{Y} = 20518 - 171(X_i)$$

The regression function does seem to be a good fit here. But observations seem to have different crime rates for the same highschool diploma percentages.

1.28b)

1)

$$\hat{Y} = 20{,}518 - 171(X)$$

$$20{,}518 - 171(81) = 6667$$

$$20{,}518 - 171(82) = 6496$$

$$171$$

2)

$$20{,}518 - 171(80) = 6838$$

3)

$$\varepsilon_{10} = Y_{10} - E\{Y_{10}\} = 7{,}993 - (20{,}518 - 171(74))$$

$$= 129$$

1.28 b)

4) $\sigma^2 = \frac{\sum(Y_i - \hat{Y}_i)^2}{n-2} = \frac{455,365,155...}{82}$

Got on F2 of HW3_1.28 spreadsheet

5,553,233.67

2.4a)

$b_1 \pm t(1-a/2; n-2) s\{b_1\}$

$b_1 = 0.039$

$t(1-0.01/2; 120-2)$

$s^2\{b_1\} = \frac{MSE}{\sum(X_i - \bar{X})^2} = \frac{44.88}{2379.925} = 0.019$

$\sqrt{0.019} = 0.14$

$MSE = \frac{\sum(Y_i - \hat{Y}_i)^2}{n-2} = \frac{0.99}{118} = 44.88$

$0.039 \pm \underbrace{t(0.995; 118)}_{2.86} \cdot 0.14 = (-0.3614 ; 0.4394)$

Note: this value comes from table B.2 from the appendix in the textbook. I used the 120 value as there was no 118 value in the table

2.4a)

the confidence interval is

$-0.3614 \leq B_1 \leq 0.4394$

with 99% confidence. This interval does include 0 & that's interesting because it means that there is no linear association between a student's average GPA & their ACT score

2.6a)

$b_1 \pm t(1-a/2; n-2) s\{b_1\} = 4 \pm \underbrace{t(0.975; 8)}_{2.306} \cdot 0.47$

$b_1 = 4$

$= 4 \pm 1.0838 \Rightarrow (-2.92; 5.08)$

$t(1-0.05/2; 10-2)$

$2.92 \leq B_1 \leq 5.08$

$s^2\{b_1\} = \frac{MSE}{\Sigma(X_i - \bar{X})^2} = \frac{2.2}{10} = 0.22$

$\rightarrow \sqrt{0.22} = 0.47$

$MSE = \frac{\Sigma(Y_i - \hat{Y}_i)^2}{n-2} = \frac{17.6}{10-2} = 2.2$

Since 0 isn't included in the confidence interval, we know that ther is a linear relationship between student's GPA & their ACT score

2.6c)

$$b_0 \pm t(1-0.05/2; 10-2)\, s\{b_0\}$$

$$s^2\{b_0\} = MSE\left[\frac{1}{n} + \frac{\bar{X}^2}{\Sigma(X_i-\bar{X})^2}\right] = 2.2\left[\frac{1}{10} + \frac{1^2}{10}\right] = 2.2(0.2)$$

$$\sqrt{0.44} = 0.66$$

$$10.2 \pm 2.306 \cdot 0.66 \Rightarrow 8.68 \leq B_0 \leq 11.72$$

Since 0 isn't included in our confidence interval, it means that B_0 does have an impact on the outcome of our linear regression model, even if $X=0$

2.13a)

$$\hat{Y}_h \pm t(1-a/2; n-2)\, s\{\hat{Y}_h\}$$

$$s^2\{\hat{Y}_h\} = MSE\left[\frac{1}{n} + \frac{(X_h-\bar{X})^2}{\Sigma(X_i-\bar{X})^2}\right] = 44.88\left[\frac{1}{120} + \frac{28-24.725}{2379.925}\right]$$

$$= 44.88 \cdot 0.0097 = 0.44 \Rightarrow \sqrt{0.44} = 0.66$$

$$3.202 \pm \underbrace{t(0.975; 118)}_{2.86} \cdot 0.66 = 1.31 \leq E\{Y_h\} \leq 5.09$$

3.13a)

We can conclude with 95% confidence that the mean GPA of a student with a 28 score on the ACT is between 1.31 & 5.09