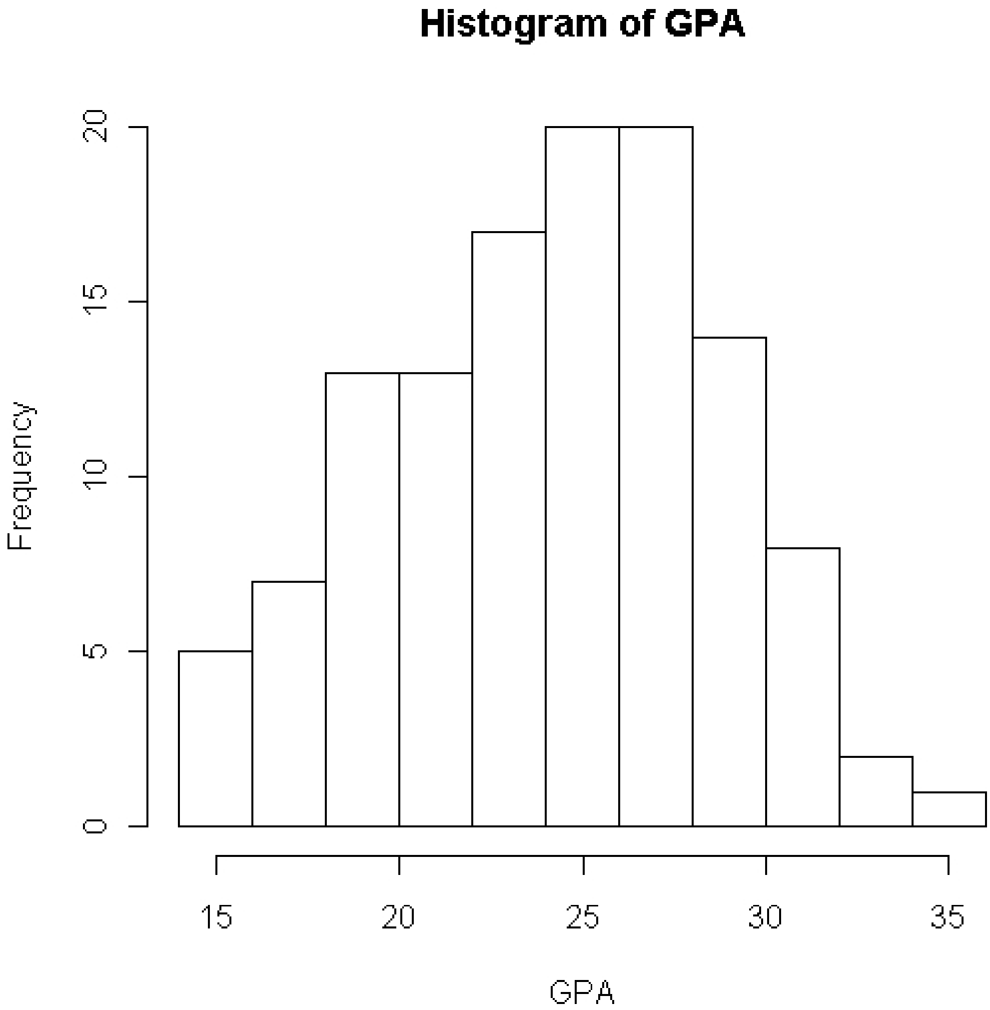
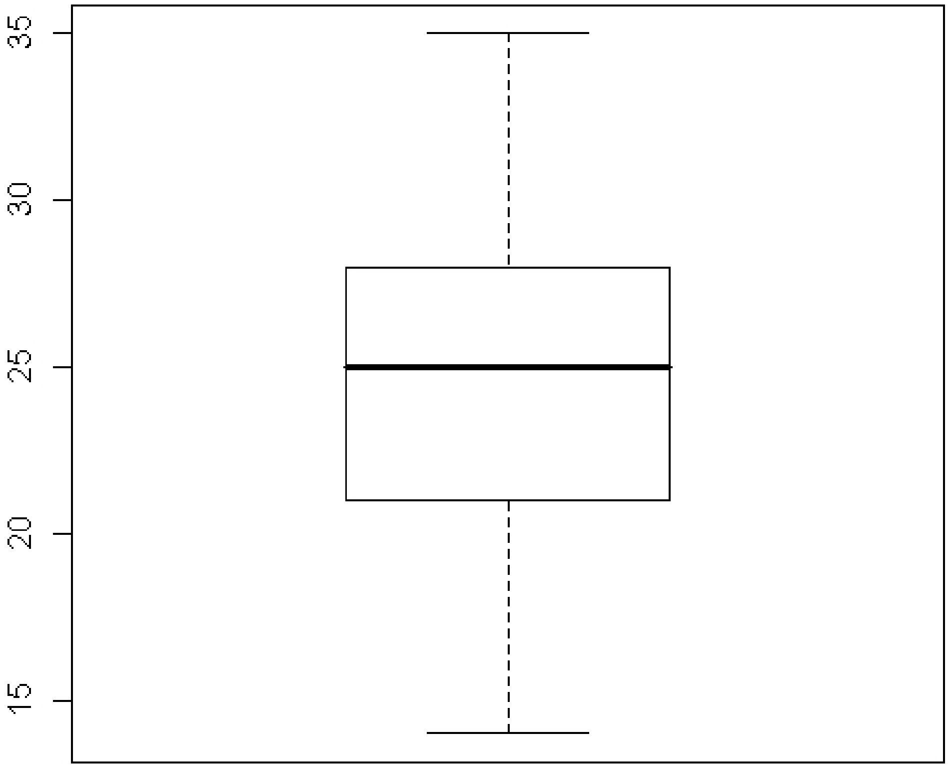
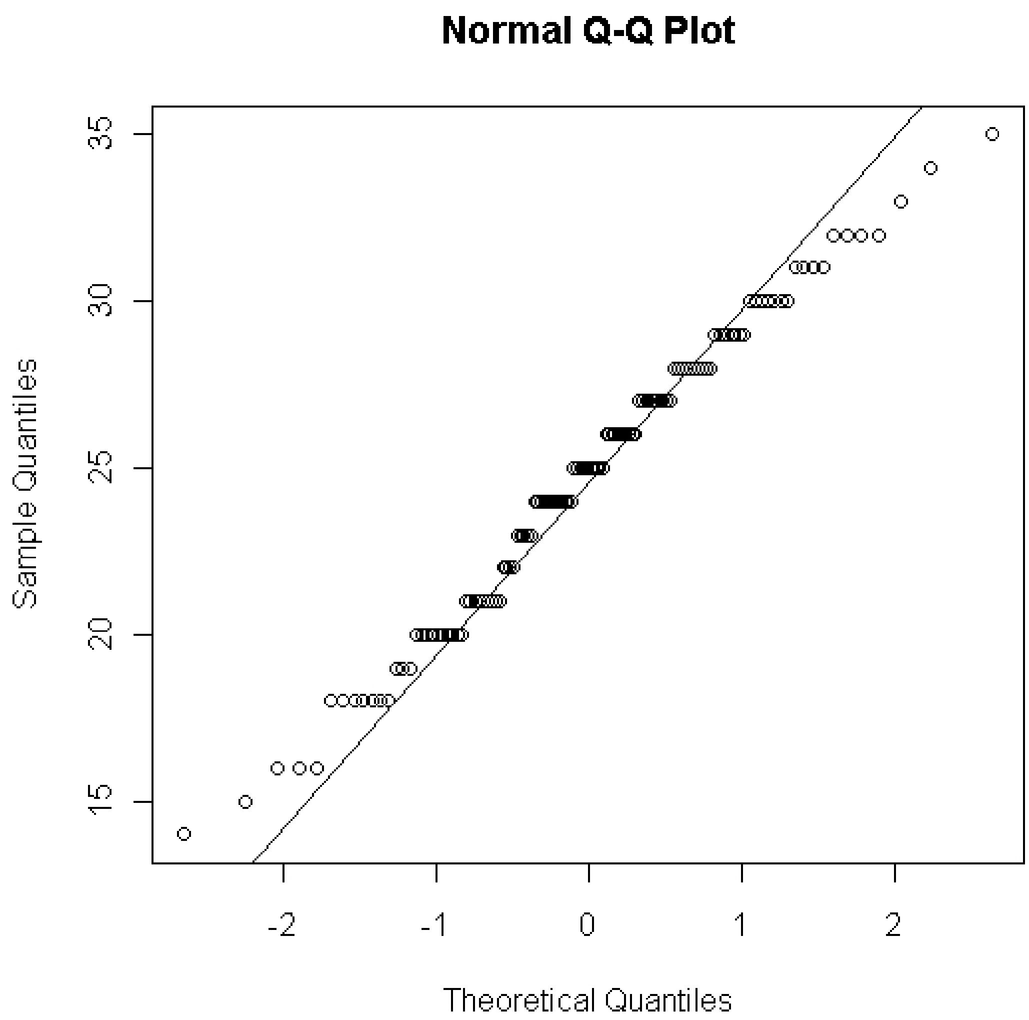
1.(a)i.



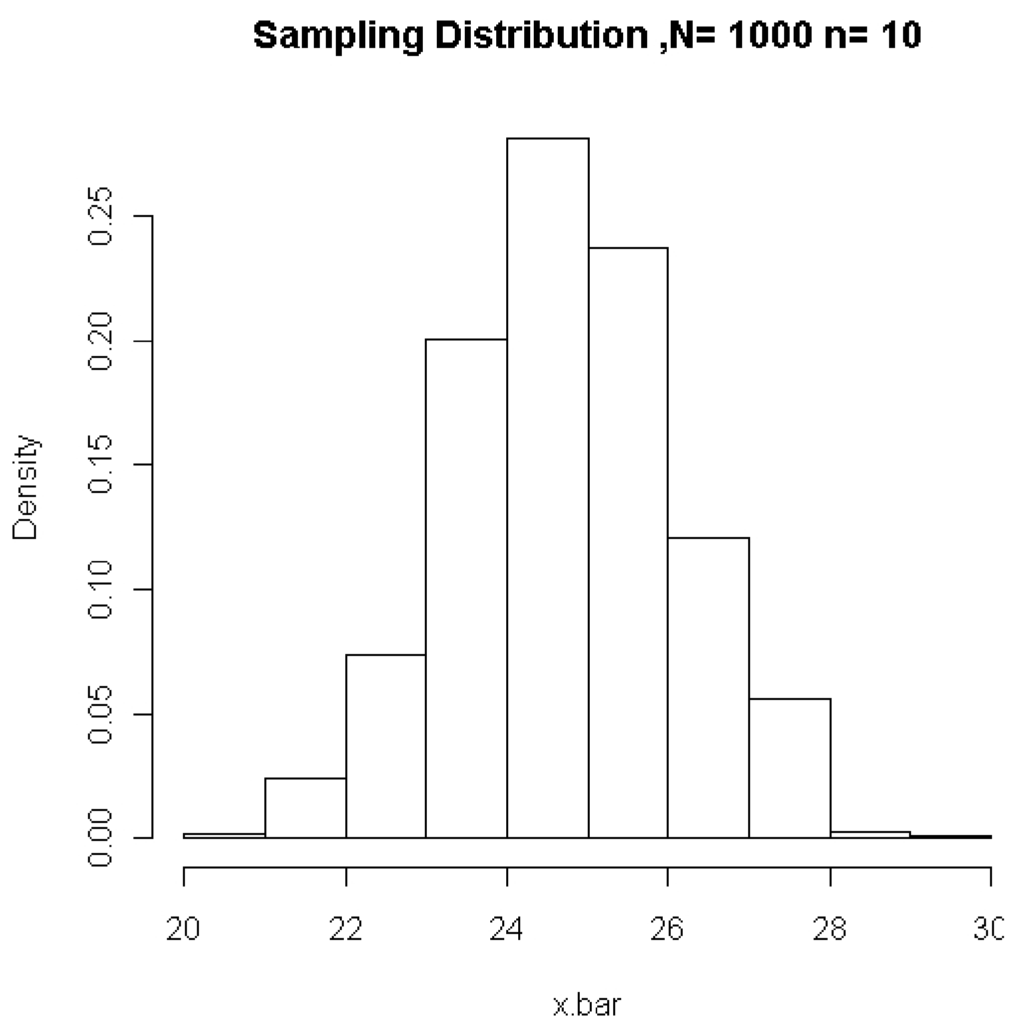
ii.



iii.


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
15
20
25
30
35
-2
-1
0
1
2

(b)



(c)令 $X =$GPA 中 major 決定的分數

$Y =$GPA 中 major 尚未決定的分數

i. $H_0 : \sigma_X^2 = \sigma_Y^2$

如果我們不 reject 這個假設，在σ未知下，我們可以考慮

$H_0 : \mu_X = \mu_Y$

ii.第一個部份

$H_1 : \sigma_X^2 \neq \sigma_Y^2$

F test to compare two variances

data: x and y
F = 0.9976, num df = 41, denom df = 77, p-value = 0.9848
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.5934336 1.7545555

sample estimates:
ratio of variances
0.9975851
$\alpha = 0.05, p-value > \alpha \Rightarrow$ do not reject H_0
所以我們可以考慮
第二個部分
$H_1 : \mu_X \neq \mu_Y$
Two Sample t-test

data: x and y
t = 1.0079, df = 118, p-value = 0.3156
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.8321791 2.5574538
sample estimates:
mean of x mean of y
25.28571 24.42308
$\alpha = 0.05, p-value > \alpha \Rightarrow$ do not reject H_0 亦即兩者並無顯著差異

2.令 $X_i = \begin{cases} 1, & \text{慣用左手的人} \\ 0, & \text{不是慣用左手的人} \end{cases}, X_i \sim Bernoulli(0.36)$

1.令 $Y = \dfrac{\sum_{i=1}^{25} X_i}{25} \Rightarrow M_Y(t) = E[\exp(t\dfrac{\sum_{i=1}^{25} X_i}{25})] = (E[e^{\frac{tX_i}{25}}])^{25} = (0.64 + 0.36e^{\frac{t}{25}})^{25}$

$\therefore p(X = \dfrac{i}{25}) = C_i^{25} 0.36^i \cdot 0.64^{25-i}, i = 0,1,2,\ldots 25$

2.令 $Z_{225} = \dfrac{\sum_{i=1}^{225} X_i}{225} \to N(0.36, \dfrac{0.36 \cdot 0.64}{225})$

$E[Z_{225}] = 0.36, V(Z_{225}) = \dfrac{0.36 \cdot 0.64}{225} = 0.001024$

3. $\sigma_{Z_{225}} = 0.032$

$P(|Z_{225} - 0.36| \leq 0.08) = P(\left|\dfrac{Z_{225} - 0.36}{0.032}\right| \leq \dfrac{0.08}{0.032}) = P(\left|\dfrac{Z_{225} - 0.36}{0.032}\right| \leq 2.5)$

$\approx P(Z \leq 2.5) = \Phi(2.5) - \Phi(-2.5) = 2\Phi(2.5) - 1 \approx 2 \cdot 0.99 - 1 = 0.98$

3.(a) $b_1 = \dfrac{\sum XY - \dfrac{\sum X \sum Y}{n}}{\sum X^2 - \dfrac{(\sum X)^2}{n}}$

$$= \frac{186659 - \dfrac{1475 \cdot 12890}{100}}{24459 - \dfrac{1475^2}{100}}$$

$$= \frac{18665900 - 1475 \cdot 12890}{2445900 - 1475^2}$$

$$= \frac{-346850}{270275} = -1.28$$

$$b_0 = \bar{Y} - b_1 \bar{X} = \frac{\sum Y}{n} - b_1 \frac{\sum X}{n} = \frac{12890}{100} + \frac{346850}{270275} \cdot \frac{1475}{100} = 147.83$$

$$\therefore \hat{Y} = b_0 + b_1 X = 147.83 - 1.28X$$

(b) $H_0 : \beta_1 = 0, H_1 : \beta_1 \neq 0$

令 $T = \dfrac{b_1 - 0}{s.e.(b_1)} \sim t(n-2)$

$$MS(E) = \frac{\sum (Y - \hat{Y})^2}{n-2} = \frac{\sum (Y - b_0 - b_1 X)^2}{98}$$

$$= \frac{1}{98} \sum (Y^2 + b_0^2 + b_1^2 X^2 - 2b_0 Y - 2b_1 XY + 2b_0 b_1 X)$$

$$= \frac{1}{98} (\sum Y^2 + 100 b_0^2 + b_1^2 \sum X^2 - 2b_0 \sum Y - 2b_1 \sum XY + 2b_0 b_1 \sum X)$$

$$= \frac{1}{98} (1714421 + 100 \cdot 147.83^2 + 1.28^2 \cdot 24459$$
$$- 2 \cdot 147.83 \cdot 12890 + 2 \cdot 1.28 \cdot 186659 - 2 \cdot 147.83 \cdot 1.28 \cdot 1475)$$

$$= \frac{1}{98} (1714421 + 2185370.89 + 40073.6256$$
$$- 3811057.4 + 477847.04 - 558206.08)$$

$$= \frac{48449.0756}{98} = 494.38$$

$$\sum (X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n} = 24459 - \frac{1475^2}{100} = 2702.75$$

$$s.e.(b_1) = \sqrt{\frac{MS(E)}{\sum (X - \bar{X})^2}} = \sqrt{\frac{494.38}{2702.75}} = 0.43$$

$$T = \frac{-1.28}{0.43} = -2.98, t_{0.975}(98) = -1.984467, T < t_{0.975}(98) \Rightarrow reject\ H_0$$

(c) $H_0 : y_{15} = \hat{y}_{15}, H_1 : y_{15} \neq \hat{y}_{15}$

令 $T = \dfrac{Y_{15} - \hat{Y}_{15}}{s.e.(Y_{15} - \hat{Y}_{15})} \sim t(n-2)$

$\hat{Y}_{15} = 147.83 - 1.28 \cdot 15 = 128.63$

$$s.e.(Y_{15} - \hat{Y}_{15}) = \sqrt{MS(E)\left(1 + \frac{1}{n} + \frac{(X_{15} - \bar{X})^2}{\sum (X - \bar{X})^2}\right)}$$

$$= \sqrt{494.38\left(1 + \frac{1}{100} + \frac{(15 - 14.75)^2}{2702.75}\right)}$$

$$= \sqrt{494.38(1 + 0.01 + 0.00002)}$$

$$= \sqrt{494.38 \cdot 1.01002} = \sqrt{499.3336876} = 22.35$$

$T = \dfrac{128.63}{22.35} = 5.76, t_{0.025}(98) = 1.984467, T > t_{0.025}(98) \Rightarrow reject\ H_0$

(d)

Source of Variantion	SS	df	MS	F
Regression	4450.9244	1	4450.9244	9.00
Error	48449.0756	98	494.38	
Total	52900	99		

$$SS(TO) = \sum (Y - \bar{Y})^2 = \sum Y^2 - \frac{(\sum Y)^2}{n} = 1714421 - \frac{12890^2}{100} = 52900$$

$$SS(R) = SS(TO) - SS(E) = 52900 - 48449.0756 = 4450.9244$$

$$F = \frac{MS(R)}{MS(E)} = \frac{4450.9244}{494.38} = 9.00$$

4.(a)令 $S(\beta_1) = \sum_{i=1}^{n} (Y_i - \beta_1 X_i)^2$

$$S'(\beta_1) = -2\sum_{i=1}^{n} X_i (Y_i - \beta_1 X_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} X_i Y_i - \beta_1 \sum_{i=1}^{n} X_1^2 = 0$$

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_1^2}$$

(b) $L(\beta_1) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_1 X_i)^2}$

$$\ln L(\beta_1) = (-\frac{n}{2})\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_1 X_i)^2$$

$$\frac{\mathrm{d}L(\beta_1)}{\mathrm{d}\beta_1} = -\frac{2X_i}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_1 X_i) = 0$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_1^2}$$

(c) $E[\widehat{\beta}_1] = \frac{1}{\sum_{i=1}^{n} X_i^2}\sum_{i=1}^{n} X_i E[Y_i] = \frac{1}{\sum_{i=1}^{n} X_i^2}\beta_1\sum_{i=1}^{n} X_i^2 = \beta_1$

$\widehat{\beta}_1$ 是 β_1 的 unbiased estimator