Chapter 8 Question 6.

Solution 1. *For part (a): let's first input the data:*

```
hour <- c(rep(12, 17), rep(18, 17), rep(24, 17), rep(30, 17))
length <- c( 5, 11,  8, 11,  4,  4,  8,  3,  6,  4,  7,  3,  5,  4,  6,  9,  3,
            11, 16, 18, 24, 18, 18, 21, 14, 21, 19, 17, 24, 14, 20, 16, 20, 22,
            17, 16, 26, 18, 14, 24, 18, 14, 24, 26, 21, 21, 22, 19, 14, 19, 19,
            20, 18, 22, 20, 21, 17, 16, 23, 25, 19, 21, 20, 27, 25, 22, 23, 23)
soak <- data.frame(hour, length)
```

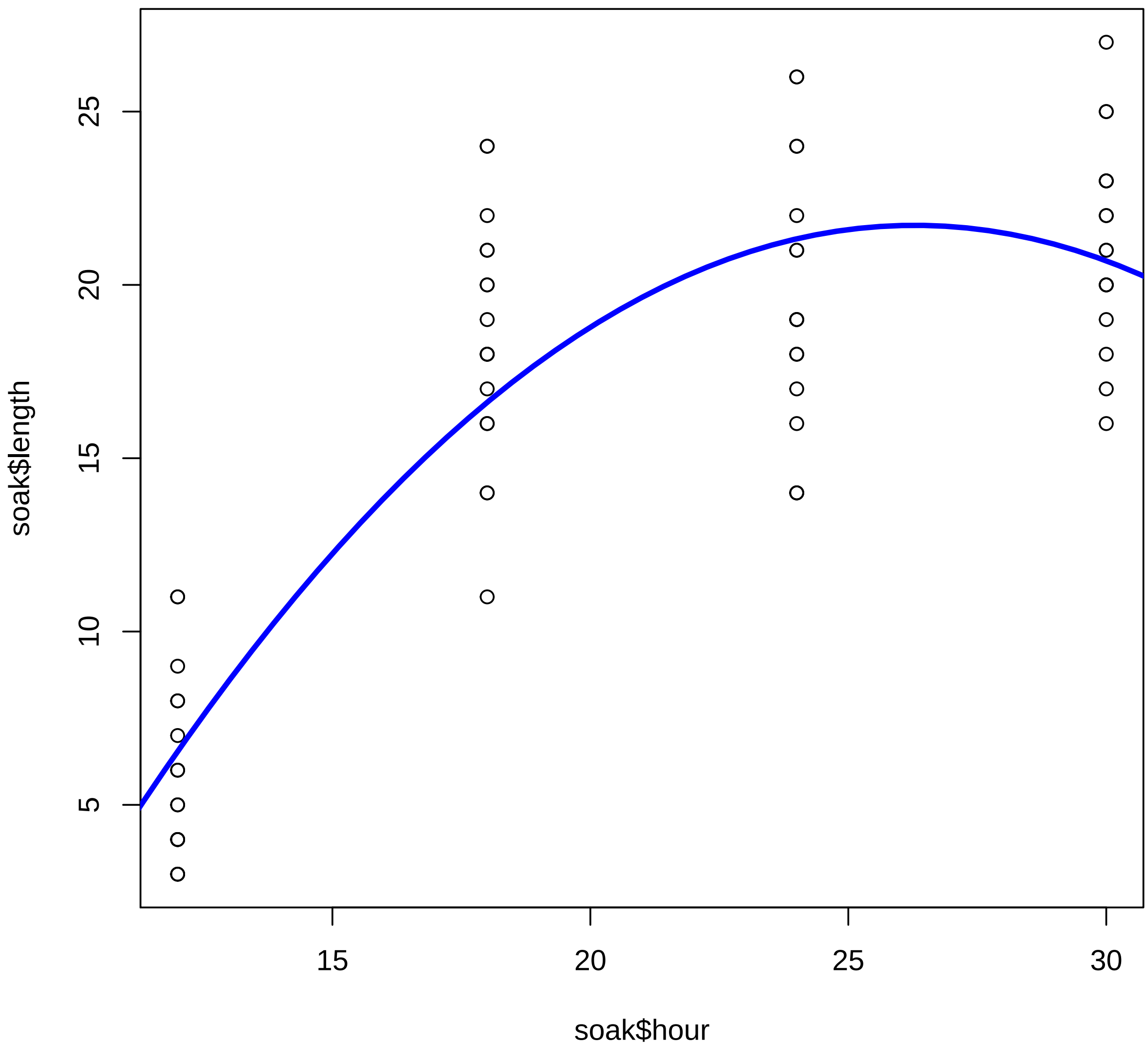
Then fit the quadratic regression model to the data:

```
soak$hour2 = soak$hour^2
fit.quad = lm(length ~ hour + hour2, data = soak)
summary(fit.quad)

##
## Call:
## lm(formula = length ~ hour + hour2, data = soak)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.3294 -2.5412 -0.3294  2.4059  7.3882
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -29.65882    4.96058  -5.979 1.05e-07 ***
## hour          3.90882    0.50820   7.691 1.03e-10 ***
## hour2        -0.07435    0.01200  -6.194 4.47e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.563 on 65 degrees of freedom
## Multiple R-squared:  0.7424,Adjusted R-squared:  0.7345
## F-statistic: 93.69 on 2 and 65 DF,  p-value: < 2.2e-16
```

then we plot the fitted response curve against the soaking time

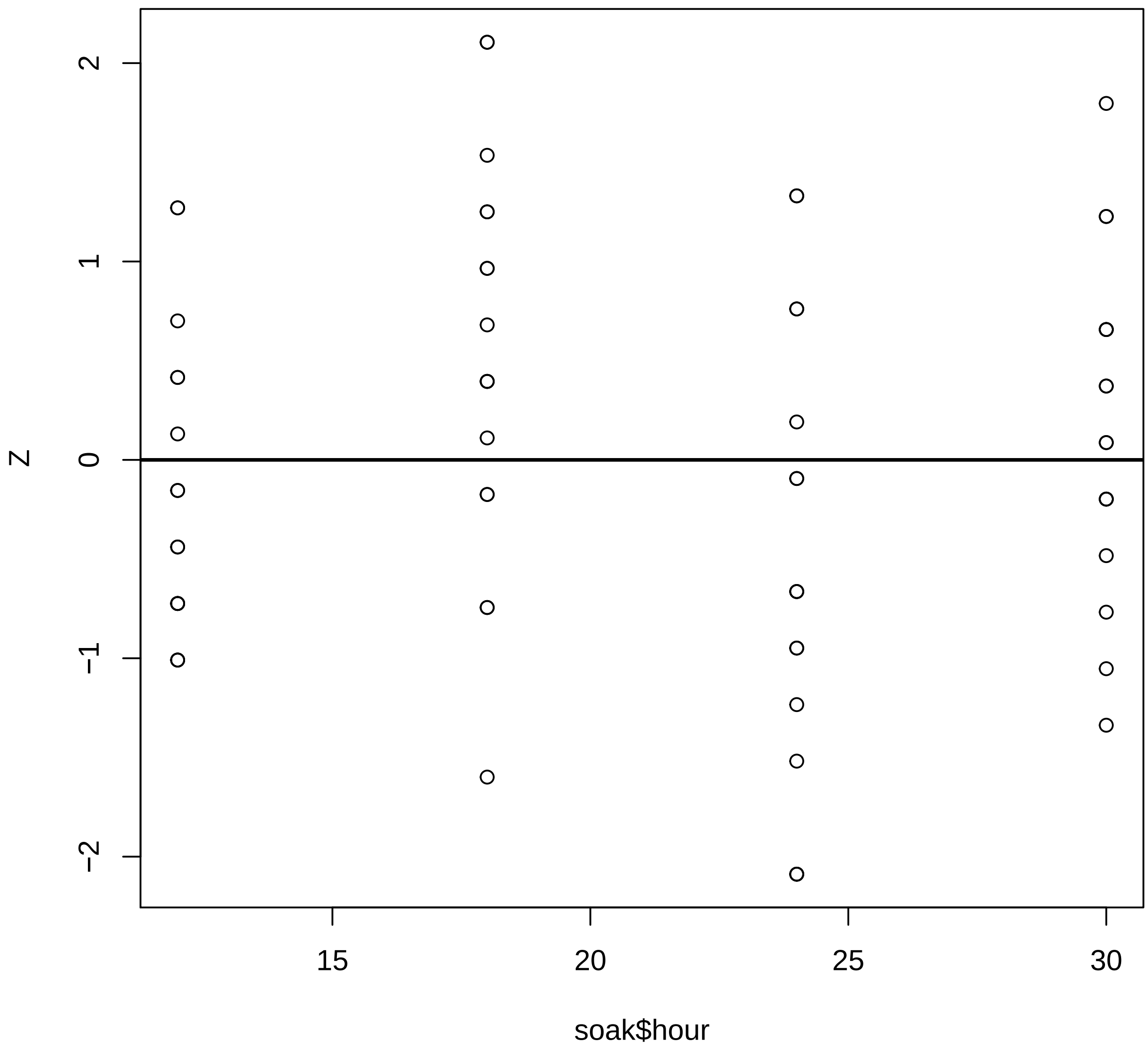
```
plot(soak$hour,soak$length)
curve(coef(fit.quad)[[1]] + coef(fit.quad)[[2]]*x + coef(fit.quad)[[3]]*x^2,
      add = T, from = 0, to = 42, lwd = 3, col = "blue")
```



It appears that at hours 18 *and* 24*, the fitted curve deviates quite far away from the mean of the reponse(length), which indicates a possibility of lack of fit.*

we can further plot the standardized residue against the soaking time:

```
SSE = sum((soak$length - fit.quad$fitted.values)^2)
Z = fit.quad$residuals/sqrt(SSE/67)
plot(soak$hour, Z)
abline(a = 0, b = 0, lwd = 2)
```



We also observe a large deviation from the within-group mean of the standardized residuals at hours 18 and 24, which indicates a possible lack of fit.

Now for part (b), let's actually run a lack of fit test.

We have null hypothesis:

$$H_0^Q : E[Y_{xt}] = \beta_0 + \beta_1 x + \beta_2 x^2$$

versus the alternative hypothesis:

$$H_A^Q : E[Y_{xt}] = \mu + \tau_x$$

We have under the alternative hypothesis:

```
#
fit.aov = aov(length ~ as.factor(hour), data = soak)
summary(fit.aov)

##                 Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(hour)  3 2501.3   833.8   75.92 <2e-16 ***
## Residuals       64  702.8    11.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

so we can compute

```
SSPE=702.8
MSPE=11.0
#sum of square and mean square for lack of fit
#degree of freedom: (N - (p + 1)) - (N - 4) = 4 - (p + 1) = 4 - 3 = 1
SSLOF = SSE - SSPE
MSLOF = SSLOF/1
MSLOF/MSPE > qf(0.95, df1 = 1, df2 = 64, lower.tail = T)

## [1] TRUE
```

According to the above lack of fit test, we reject the null hypothesis, which interprets as the quadratic polynomial regression model is NOT adequate.