Solution to HW #6

4.14
(a). Use R to generate linear regression model without interception:

```r
> gpa <- read.table("CH01PR19.txt")
> colnames(gpa) <- c("Y","X")
> lm.gpa <- lm(Y~X-1,data=gpa)
> summary(lm.gpa)
```

```
Call:
  lm(formula = Y ~ X - 1, data = gpa)

Residuals:
     Min      1Q  Median      3Q     Max
-3.0276 -0.2737  0.1077  0.4754  2.1820

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
X       0.121643   0.002637   46.13   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7257 on 119 degrees of freedom
Multiple R-squared:  0.947,  Adjusted R-squared:  0.9466
F-statistic: 2128 on 1 and 119 DF,  p-value: < 2.2e-16
```

Therefore, the coefficient for \( X \) is 0.121643.

The estimated regression function is \( \hat{Y} = 0.1216 \times X \).

(b)

```r
> confint(lm.gpa,level=0.95)
```

```
     2.5 %   97.5 %
X  0.1164216 0.1268643
```

The estimated confidence interval for \( \beta_1 \) is (0.1164,0.1268). We are 95% confident that (0.1164,0.1268) covers the true value of \( \beta_1 \).

(c)

```r
> predict.lm(lm.gpa,newdata=data.frame(X=30),interval="confidence",level=0.95)
```

```
     fit     lwr      upr
1 3.649287 3.492647 3.805928
```

The mean freshman GPA for students with ACT score 30 is 3.6493, and the 95% confidence interval is (3.4926,3.8059).
4.15  
(a)  
> plot(Y~X,data=gpa,xlim=c(0,35),ylim=c(0,4.0),main="The fitted regression line and the original data")  
> abline(lm.gpa)

The regression function through the origin is not a good fit here as the points do not spread randomly on the two sides of the line.

(b)  
> ei <- lm.gpa$res  
> sum(ei)  
[1] 7.9715  
> plot(ei~predict(lm.gpa),ylab="residual",xlab="fitted value",main="Residuals agains the fitted value")  
> abline(h=0,lty=2)
The residuals do not sum to 0. We see from the plot that residuals take more positive values when \( \hat{Y} \) is small compared with when \( \hat{Y} \) is large. There is a decreasing linear trend between residuals and \( \hat{Y} \).

(c)

\( H_0 \) (Reduced model): \( Y = \beta_1 X + \)

\( H_a \) (Full model): \( Y = \beta_0 + \beta_1 X + \)

\[
\text{lm.gpa2} \leftarrow \text{lm}(Y \sim X, \text{data=gpa})
\]
\[
\text{anova(lm.gpa,lm.gpa2)}
\]

Analysis of Variance Table

<table>
<thead>
<tr>
<th>Model 1: Y ~ X - 1</th>
<th>Model 2: Y ~ X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Res.Df</td>
<td>RSS</td>
</tr>
<tr>
<td>1</td>
<td>119</td>
</tr>
<tr>
<td>2</td>
<td>118</td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The lack of fit test F-value=43.4 (read from R output) and the corresponding P-value is 1.304*10^{-9}. Since the P-value is smaller than \( \alpha=0.005 \), we reject the null hypothesis and conclude that regression model with intercept is more appropriate compared with the model through origin.

4.22

\[
P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) = 1 - P(A_1 \cup A_2 \cup A_3) \geq 1 - (P(A_1) + P(A_2) + P(A_3)) = 1 - 3\alpha
\]