Solution to SLR Practice Problem

(a) Describe the association between number of cans of beer and BAC.
Positive linear association.

(b) What are the explanatory and response variables for the linear regression model?
Explanatory variable: Number of cans of beer
Response variable: BAC

(c) Write the equation for the least squares line.
$\hat{y} = -0.013 + 0.018x$

(d) Interpret the slope and the intercept in context.
Slope: An increase in the number of cans of beer by 1 is associated with an increase in BAC by 0.018.
Intercept: The predicted BAC for someone who had 0 cans of beer is -0.013. The regression summary also indicates that the intercept term is not significantly different than 0 (see part j).

(e) What is the predicted BAC for a person that drank 5 cans of beer?
$\hat{y} = -0.013 + 0.018(5) = 0.077$

(f) A student in this data set drank 9 beers and had a measured BAC of 0.19. Calculate the residual for this student.
$e_i = y_i - \hat{y}_i = 0.19 - [-0.013 + 0.018(9)] = 0.19 - 0.149 = 0.041$

(g) Interpret the coefficient of determination (R^2).
$R^2 = 0.7998$, which means that 79.98% of the variation in BAC can be explained by the number of cans of beer the student drank (x).

(h) Do the data provide strong evidence that drinking more cans of beer is associated with an increase in blood alcohol content? State the null and alternative hypotheses, report the test statistic and p-value (from the `summary()` command), and state your conclusion.
$H_0 : \beta_1 = 0$
$H_A : \beta_1 \neq 0$
The test statistic is $t = 7.48$ with a p-value < 0.001. Therefore, we reject H_0, and conclude that drinking more cans of beer is associated with an increase in BAC.

(i) Calculate a 95% confidence interval for β_1.
The critical value is given by $t_{0.025;n-2}$ = `abs(qt(0.025, 14))` = 2.14.

$$0.018 \pm 2.14(0.0024) \implies (0.013, 0.023)$$

We are 95% confident that β_1 is between 0.013 and 0.023.
In R:

```r
library(openintro)
lm1 <- lm(BAC ~ Beers, data = bac)
confint(lm1)

##                   2.5 %     97.5 %
## (Intercept) -0.03980535 0.01440414
## Beers        0.01281262 0.02311490
```

(j) Do the data provide evidence that the intercept is significantly different than 0? State the null and alternative hypotheses, report the test statistic and p-value (from the `summary()` command), and state your conclusion.
$H_0 : \beta_0 = 0$
$H_A : \beta_0 \neq 0$
The test statistic is $t = -1.005$ with a p-value $= 0.332 > 0.05$. Therefore, we do not reject H_0, and conclude that the intercept is not significantly different than 0. This makes sense in context since it is reasonable to assume that a person that drank 0 beers has a BAC of 0.

(k) Calculate a 95% confidence interval for β_0.

$$-0.013 \pm 2.14(0.0126) \implies (-0.04, 0.014)$$

We are 95% confident that β_0 is between -0.04 and 0.014. Note that 0 is in the interval, which agrees with the hypothesis test.

(l) Are the conditions for linear regression reasonably satisfied? In your assessment, comment on the plot of the residuals versus number of cans of beer (x), and the QQ plot of the residuals shown below.
Yes. The trend is linear. The points in the plot of the residuals versus number of beers (x) look randomly scattered, and evenly distributed around 0 (constant variance). The QQ plot indicates that the residuals are approximately normally distributed. We can assume independence since the number of beers were assigned randomly to participants.