STAT 632, Lecture 7 Handout

The Child Health and Development Studies investigate a range of topics. One study considered all pregnancies between 1960 and 1967 among women in the Kaiser Foundation Health Plan in the San Francisco East Bay area. The response variable of interest is `bwt`, the birthweight of the infants in ounces. The explanatory variables are `gestation`, the length of gestation, in days; `parity`, a dummy variable which is 0 if the child is first born, and 1 otherwise; `age`, the mother's age in years; `height`, the mother's height in inches; `weight`, the mother's weight in pounds; and `smoke`, a dummy variable which is 0 if the mother is a nonsmoker, and 1 if the mother smokes. A regression summary from fitting this model in R is shown below.

```
library(openintro)
head(babies, n=3)
```

```
##   case bwt gestation parity age height weight smoke
## 1    1 120       284      0  27     62    100     0
## 2    2 113       282      0  33     64    135     0
## 3    3 128       279      0  28     64    115     1
```

```
lm1 <-  lm(bwt ~ gestation + parity + age + height + weight + smoke, data = babies)
summary(lm1)
```

```
##
## Call:
## lm(formula = bwt ~ gestation + parity + age + height + weight +
##     smoke, data = babies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -57.613 -10.189  -0.135   9.683  51.713
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -80.41085   14.34657  -5.605 2.60e-08 ***
## gestation     0.44398    0.02910  15.258  < 2e-16 ***
## parity       -3.32720    1.12895  -2.947  0.00327 **
## age          -0.00895    0.08582  -0.104  0.91696
## height        1.15402    0.20502   5.629 2.27e-08 ***
## weight        0.05017    0.02524   1.987  0.04711 *
## smoke        -8.40073    0.95382  -8.807  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.83 on 1167 degrees of freedom
##   (62 observations deleted due to missingness)
## Multiple R-squared:  0.258,Adjusted R-squared:  0.2541
## F-statistic: 67.61 on 6 and 1167 DF,  p-value: < 2.2e-16
```

(a) Write the equation for the multiple linear regression model.

(b) Interpret the coefficients for `gestation` and `smoke`.

(c) For which predictor(s) do we reject the null hypothesis $H_0 : \beta_i = 0$?

(d) Calculate a 95% confidence interval for the coefficient of `gestation`. Note that $n = 1174$.

(e) Interpret the coefficient of determination (R^2)?

(f) The following is a coefficient table for a simple linear regression model with `bwt` as the response and `height` as the predictor. Why is the coefficient for `height` presented below different than the multiple linear regression model?

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  27.6810    13.0298   2.124   0.0338 *
height        1.4334     0.2033   7.052 2.97e-12 ***
```