hw8

Kyu Park

2021 3 7

1.

a)

```
data = na.omit(babies)
modg = lm(bwt~gestation, data = data)
summary(modg)
```

```
##
## Call:
## lm(formula = bwt ~ gestation, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -49.348 -11.065   0.218  10.101  57.704
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10.75414    8.53693   -1.26    0.208
## gestation     0.46656    0.03054   15.28   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.74 on 1172 degrees of freedom
## Multiple R-squared:  0.1661, Adjusted R-squared:  0.1654
## F-statistic: 233.4 on 1 and 1172 DF,  p-value: < 2.2e-16
```

i)

$$SE = 16.74\sqrt{1 + \frac{1}{1174} + \frac{(285 - 279)^2}{300687}} = 16.74812$$

```
qt(0.025, 1173)
```

```
## [1] -1.961988
```

$$CI = -10.7541 + (0.4666)(285) \pm (1.961988)(16.74812) = (89.36728, 155.08651)$$

(89.36728,155.08651)

ii)

$$RSS = \sqrt{\frac{1}{1172}SSE} = 16.74, SSE = 328426.7472$$

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = \frac{SST - 328426.7472}{SST} = 0.1661$$

$$SST = 393844.28252$$

b) i) Let H_o: $\beta = 0$ for all parameters and there is no correlation of baby's weight and whether mother smokes or not H_a: $\beta \neq 0$ for any parameter, and there is some correlation of baby's weight and whether mother smokes or not

ii) Since P<2e-16, at 95% confidence level, reject H_o; there is evidence to say that there is correlation of baby's weight and mother smoking.

iii) $df = 1, F = \frac{\frac{328608-309075}{1}}{\frac{309075}{1171}} = 74.01$

iv) Red dots are the data points of non-smoking mothers and green dots are the data points of smoking mothers. Solid line (regression equation for non-smoking mothers): $y = 0.45x - 3.18$ Dotted line (regression equation for smoking mothers): $y = 0.45x - 3.18 - 8.37 = 0.45x - 11.55$ for y = baby weight and x = gestation.

c)

```
modfull = lm(bwt~gestation+smoke+parity+age+height+weight, data = data)
summary(modfull)$coefficients #coefficients
```

```
##                  Estimate  Std. Error    t value     Pr(>|t|)
## (Intercept) -80.410853396 14.34656939 -5.6048837 2.598856e-08
## gestation     0.443978339  0.02909777 15.2581590 4.573771e-48
## smoke        -8.400733484  0.95382073 -8.8074554 4.543169e-18
## parity       -3.327199613  1.12894913 -2.9471652 3.270767e-03
## age          -0.008950305  0.08581984 -0.1042918 9.169557e-01
## height        1.154020364  0.20501847  5.6288606 2.270545e-08
## weight        0.050165027  0.02524155  1.9873987 4.711167e-02
```

```
summary(modfull)$sigma #residual standard error
```

```
## [1] 15.82924
```

```
summary(modfull)$df #degrees of freedom
```

```
## [1]    7 1167    7
```

i) $y = -80.41 + 0.44g - 8.40s - 3.33p - 0.01a + 1.15h + 0.05w$ where g = gestation, s = smoke, p = parity, a = age, h = height, w = weight.

ii) For every 1 unit increase in gestation, the weight increases by 0.44 unit. For every 1 unit increase in smoke, the weight decreases by 8.40 unit.

iii)

$$SSE = (RSE^2)(df) = 1167(15.83^2) = 292437$$

$$SST = SSR + SSE = 292437 + 101649 = 394086$$

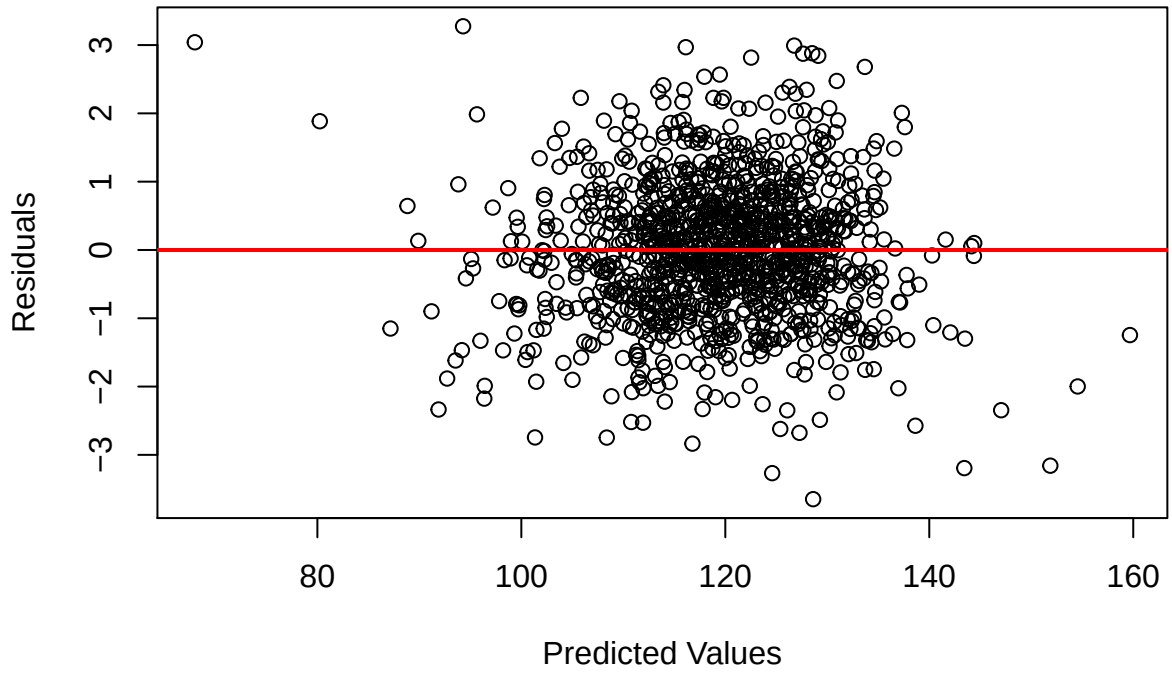
iv)

$$R^2 = \frac{SSR}{SST} = \frac{101649}{394086} = 0.2579$$

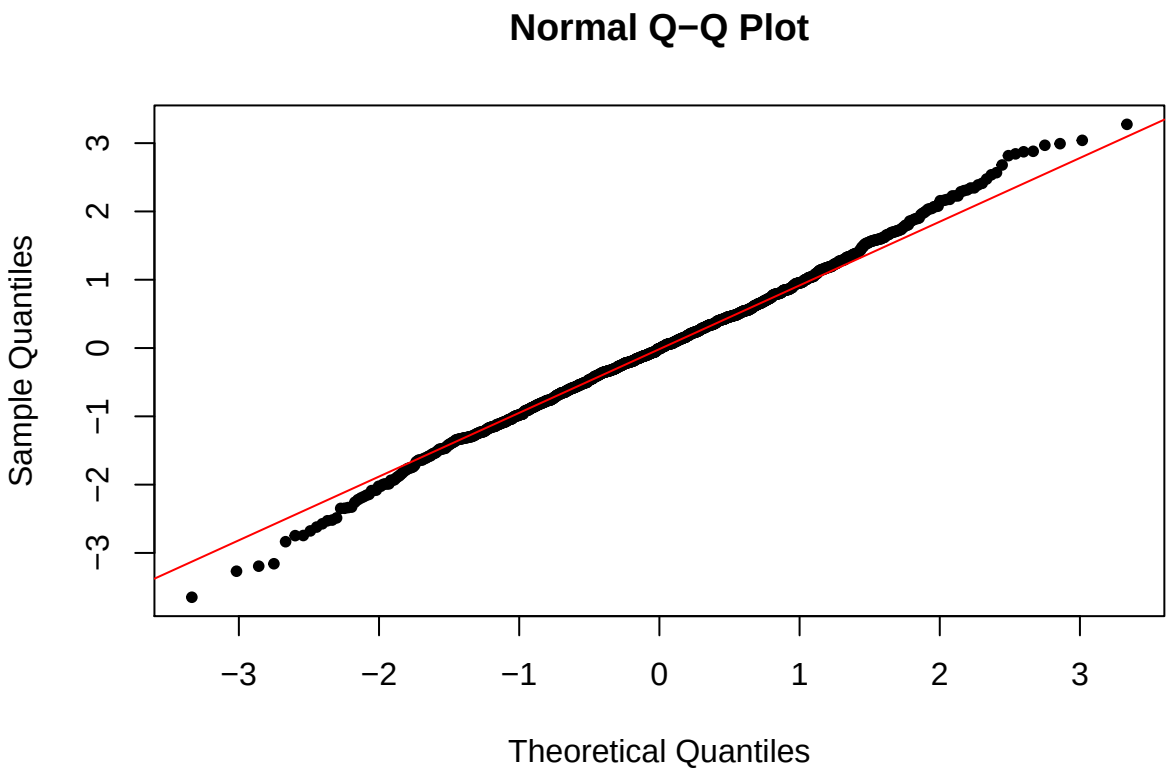
$$Adjusted R^2 = 1 - \frac{SSE/(n-7)}{SST/(n-1)} = 1 - \frac{292437/1167}{394086/1173} = 0.25412$$

v)

```
res = (modfull$res - mean(modfull$res))/sd(modfull$res)
plot(modfull$fitted.values, res, xlab = "Predicted Values", ylab = "Residuals")
abline(h=0, col = "RED", lwd = 2)
```




```
qqnorm(res, pch = 20)
qqline(res, col = "red")
```



Model assumptions are satisfied, as they are normally distributed, have constant variance around 0 and show no clear pattern with no clear outlier.

d)

```
modpart = lm(bwt~gestation+smoke+parity+height, data = data)
anova(modpart, modfull)
```

```
## Analysis of Variance Table
##
## Model 1: bwt ~ gestation + smoke + parity + height
## Model 2: bwt ~ gestation + smoke + parity + age + height + weight
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1   1169 293404
## 2   1167 292409  2     995.1 1.9857 0.1377
```

i) H_o: The coefficients of age and weight for baby's weight are 0; no correlation of age and weight to baby's weight H_a: The coefficients of age and weight for baby's weight are not 0; there is correlation of age and weight to baby's weight

ii) Since Pr(>F) = 0.1377, at 95% confidence level, accept null hypothesis; there is no evidence to say that there is correlation of age and weight to body weight.