STAT 217: Quiz 20

A high respiratory rate is a potential diagnostic indicator of respiratory infection in children. To judge whether a respiratory rate is truly high, a physician must have a clear picture of the distribution of normal respiratory rates. To this end, Italian researchers measured the respiratory rates of 618 children between the ages of 15 days and 3 years.

```
require(Sleuth2)
kids<-ex0824
lm.kids <- lm(Rate~Age, data=kids)
summary(lm.kids)

##
## Call:
## lm(formula = Rate ~ Age, data = kids)
##
## Residuals:
##    Min     1Q  Median     3Q    Max
## -19.65  -5.43  -0.61   4.59  32.27
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  47.0522     0.5042    93.3   <2e-16 ***
## Age          -0.6957     0.0294   -23.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.84 on 616 degrees of freedom
## Multiple R-squared:  0.477,Adjusted R-squared:  0.476
## F-statistic:  561 on 1 and 616 DF,  p-value: <2e-16
```
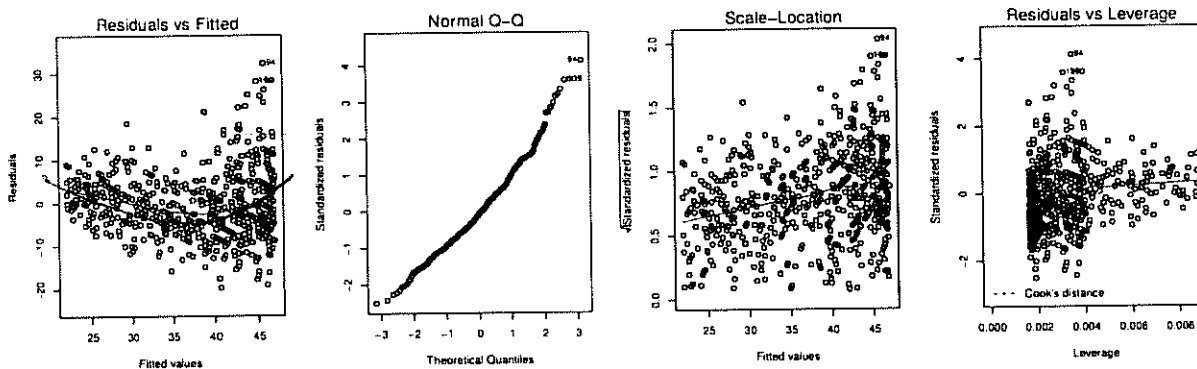
1. Circle the coefficient of determination in the above output.

2. Choose the correct interpretation of R^2 in this context.

   A. 47.66% of the variation in children's ages is explained by the linear model with respiratory rates as a predictor.

   B. 47.66% of the variation in children's respiratory rates is explained by the linear model with age as a predictor.

   C. There is a moderate relationship between age and respiratory rate.

   D. The coefficient of determination is a measure of the strength and direction of a linear relationship.

1

3. For each of the following assumptions and conditions for a simple linear regression model, say whether it is met or not. Explain how you know. Use the plots below.

```
plot(kids, main="Plot of All the Data")
```


Plot of All the Data
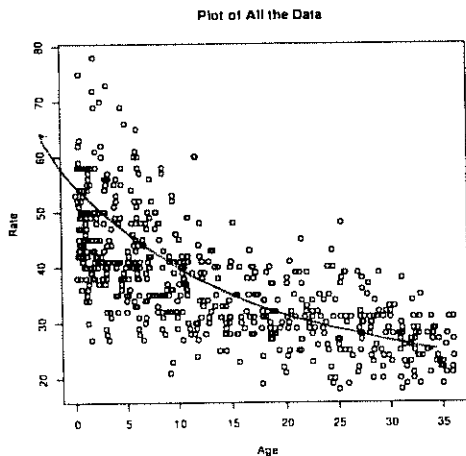
```
par(mfrow=c(1,4))
plot(lm.kids)
```


Residuals vs Fitted   Normal Q-Q   Scale-Location   Residuals vs Leverage

Quantitative variables condition: *Yes - Age + Rate are both quantitative*

Independence Assumption: *For this one, just speculate on one possible violation.*

If multiple children within one family were sampled, we might see similarities in the respiratory rates among children within families.

2

Linearity: Does not appear to be met. The scatterplot and residuals vs. fitted values plot has a curved pattern.

Constant Variance Assumption:

Does not appear to be met. The spread is greater for higher fitted values. Also, we see an upward trend in the scale-location plot.

Normality: Appears to be adequately met although there is a long right tail. This suggests that the distribution of respiratory rates is right skewed.

No influential points:

Is met. There are no points flagged on the residuals vs. leverage plot.