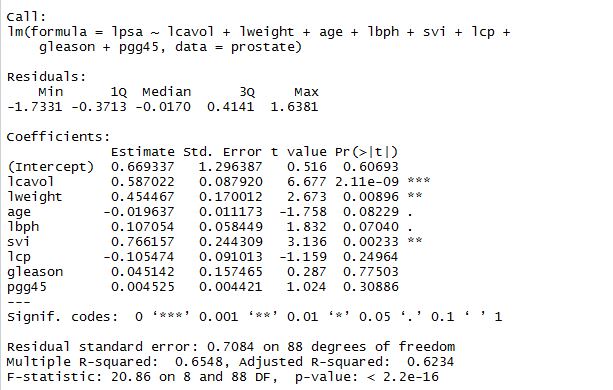
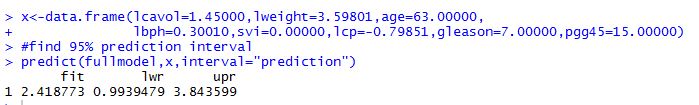
**Problem 2**

**The full model:**



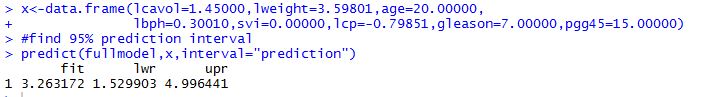
1. **Prediction: 2.419**

**95% PI: (0.994, 3.844)**



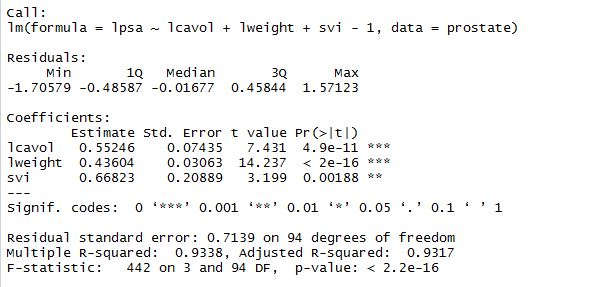
1. **Prediction: 3.263**

**95% PI: (1.530, 5.000)**



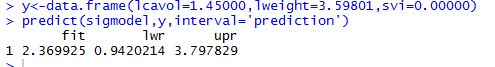
Obtain the mean vector for x, and as shown in the picture above, 63 is a lot closer to the mean of age than 20, so when change age from 63 to 20, the standard error gets larger, and therefore we obtain a wider prediction interval.

1. **The reduced model: (variables kept: “lcavol”, “lweight”, and “svi”.**



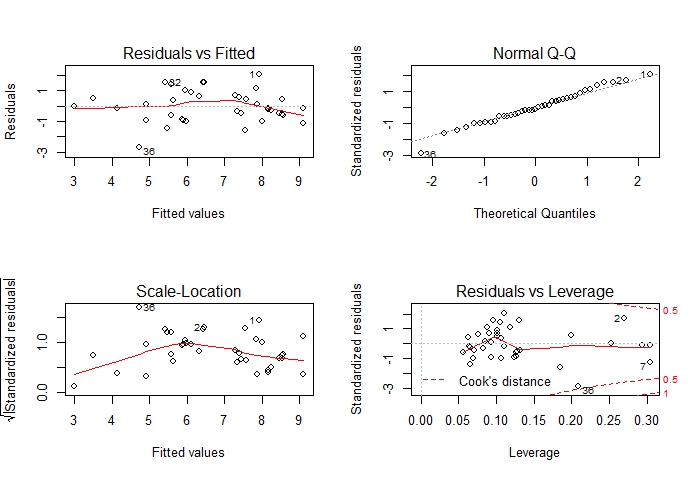
**Prediction: 2.370**

**95% PI: (0.942, 3.798)**

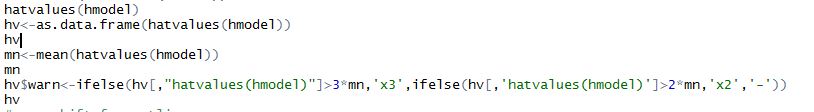


The prediction interval for the reduced model is wider than that in (a) but narrower than that in (b). I would prefer predictions obtained via the full model, because even though the variable age is not significant at α = 0.05, the p-value is still relatively small (0.1). Considering the information this variable could contain, I believe it would be better if we make predictions with age serving as a predictor.

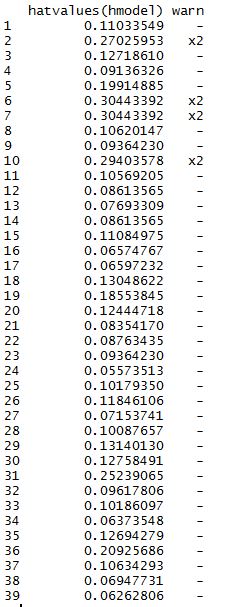
**Problem 3**



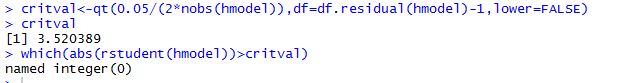
1. By looking at the left-bottom plot (Scale-Location), we can see that the square root of the absolute value of residuals does not maintain well a flat pattern, especially at the lower end, indicating some problems with the constant variance assumption. Cases 1, 2, and 36 may be worth taking a look at.
2. According to the Q-Q plot (top-right corner), we can see that normality is met by most part, except for the lower end, where several observations are off the straight line. Case 36 is way off the line, so we can consider removing this one.
3. Using the code below, we can obtain the cases whose leverages are twice or three times as large as the mean leverage,



And here’s what we got, and turns out Cases 2, 6, 7, and 10 have large leverages, and they are larger than twice the mean leverage, so we might want to further check them.

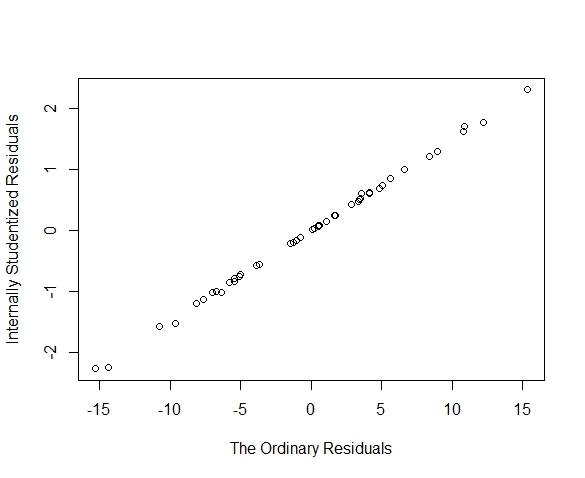
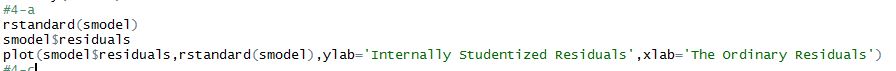


1. According to the mean shift test (code shown below), we can say that there are there’s no outliers.

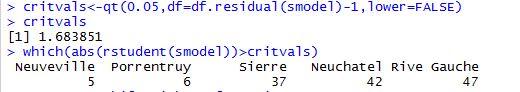
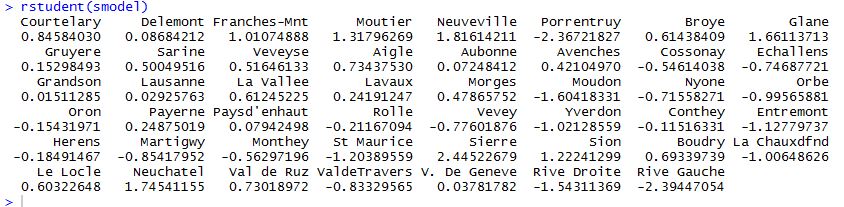
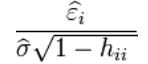
****

1. According to the Residuals vs. Leverage plot (right-bottom corner), we can see that most of the cases are fine, not exceeding either Cook’s D=1. We can conclude that there’re no influential points.

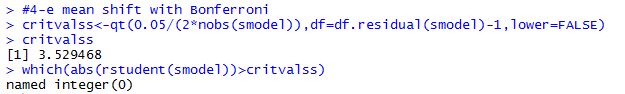
**Problem 4**



1. The formula for internally studentized residuals is , and because not all points have the same standard error (the denominator in the formula), the points do not fall on one straight line.



Outliers: Neuveville, Porrentruy, Sierre, Newuchatel, and Rive Gauche.



No outliers.