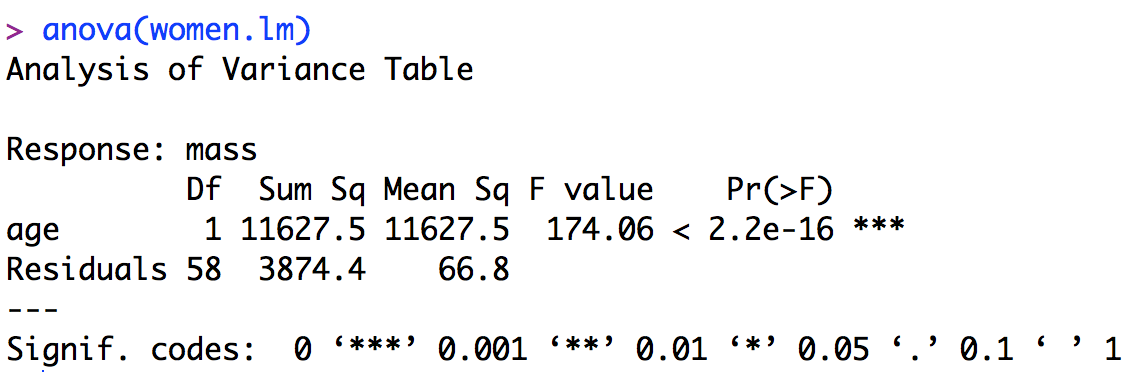
**Stats 201 HW3**

2.18

We can do one-sided test in t test, but not in F test. Therefore, t test can do better if we are sure about the side we are testing.

2.29

b.



c.

1) Alternative hypothesis: β1 ≠ 0.

2) Decision rule: if F\* > F(0.95;1, 58) = 2.37, reject β0.

3) Conclusion: because F\* = 174.06 > 2.37, reject β0, which means that there is a linear relationship between age and muscle mass.

d.



The proportion of unexplained error is relative small than the error explained by regression.

e.



Since the slope is negative, r = 

2.33

b.

Full model: Yi = β0 + β1Xi + εi

Reduced model: Yi = 7.5 + β1Xi + εi

c. dfF is still n-2 as we have discussed in class. dfR is n-1 since now we only have to estimate β1, we only lose one degree of freedom from n. Therefore, dfR - dfF = 1.

2.56a

E{MSE} = σ2 = (0.6)2 = 0.36

E{MSR} = σ2 + β12 \* SSX = 0.36 + 32 \* 114 = 1026.36

[SSX = (1-8)2 + (4-8)2 + (10-8)2 + (11-8)2 + (14-8)2 = 114]

2.57a

The reduce model is Yi = β0 + 5Xi + εi, and since now we only have to estimate β0, we only lose one degree of freedom. Therefore, dfR = n-1.

3.1

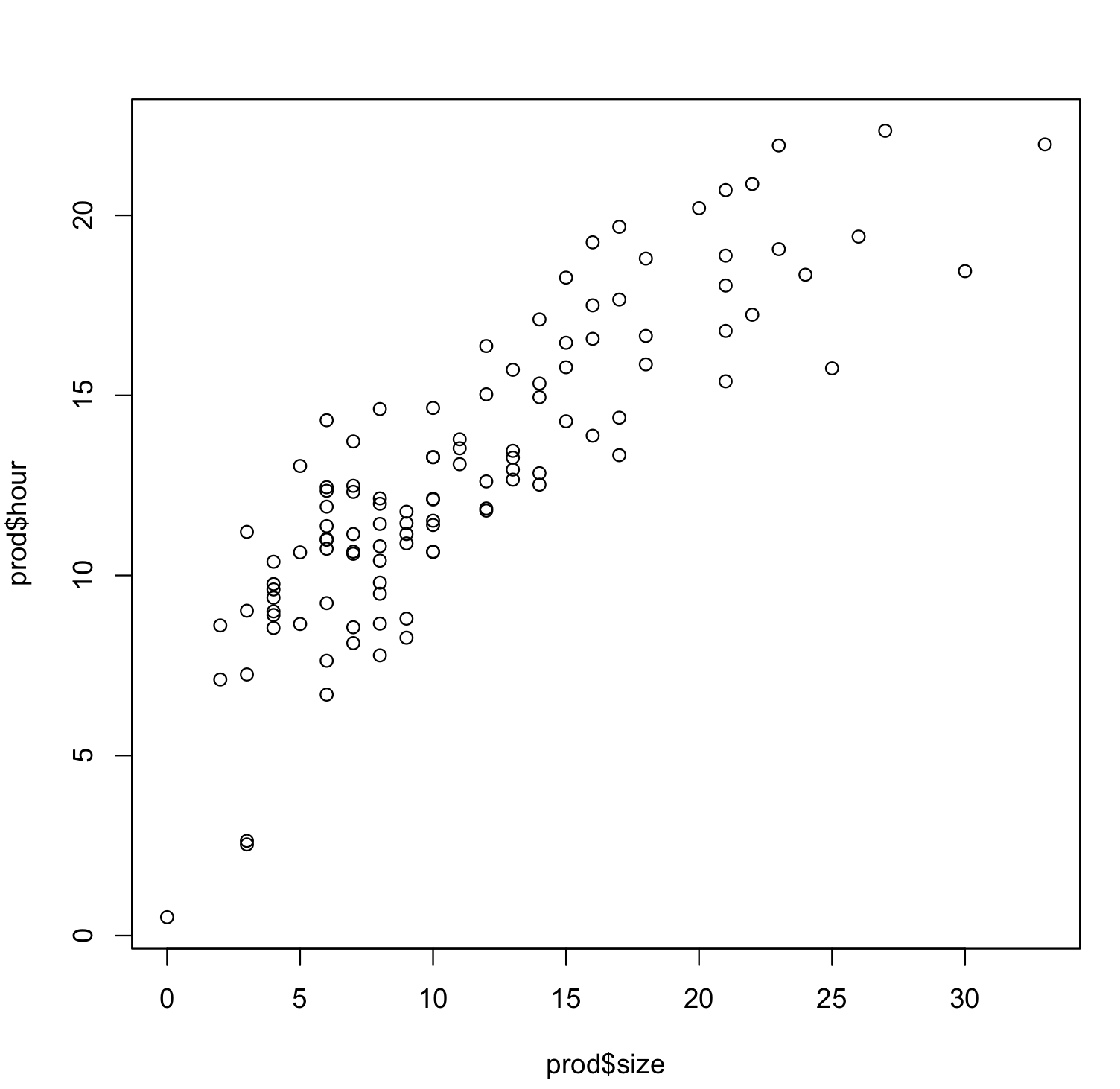
(1) Residual is the difference between the observed Y value and the predict Y value using our estimate of the regression model. Semistudentized residual is the “standardized” residual and is given by .

(2) E{εi} = 0 means that the expected value of the true errors εi = 0 because of the normal error assumption we made, but ē = 0 means that the average of residuals is zero (which is always the case in our regression model).

(3) The error term is the true error in the regression model, while the residual is the difference between the observed Y value and the predicted Y value.

3.18

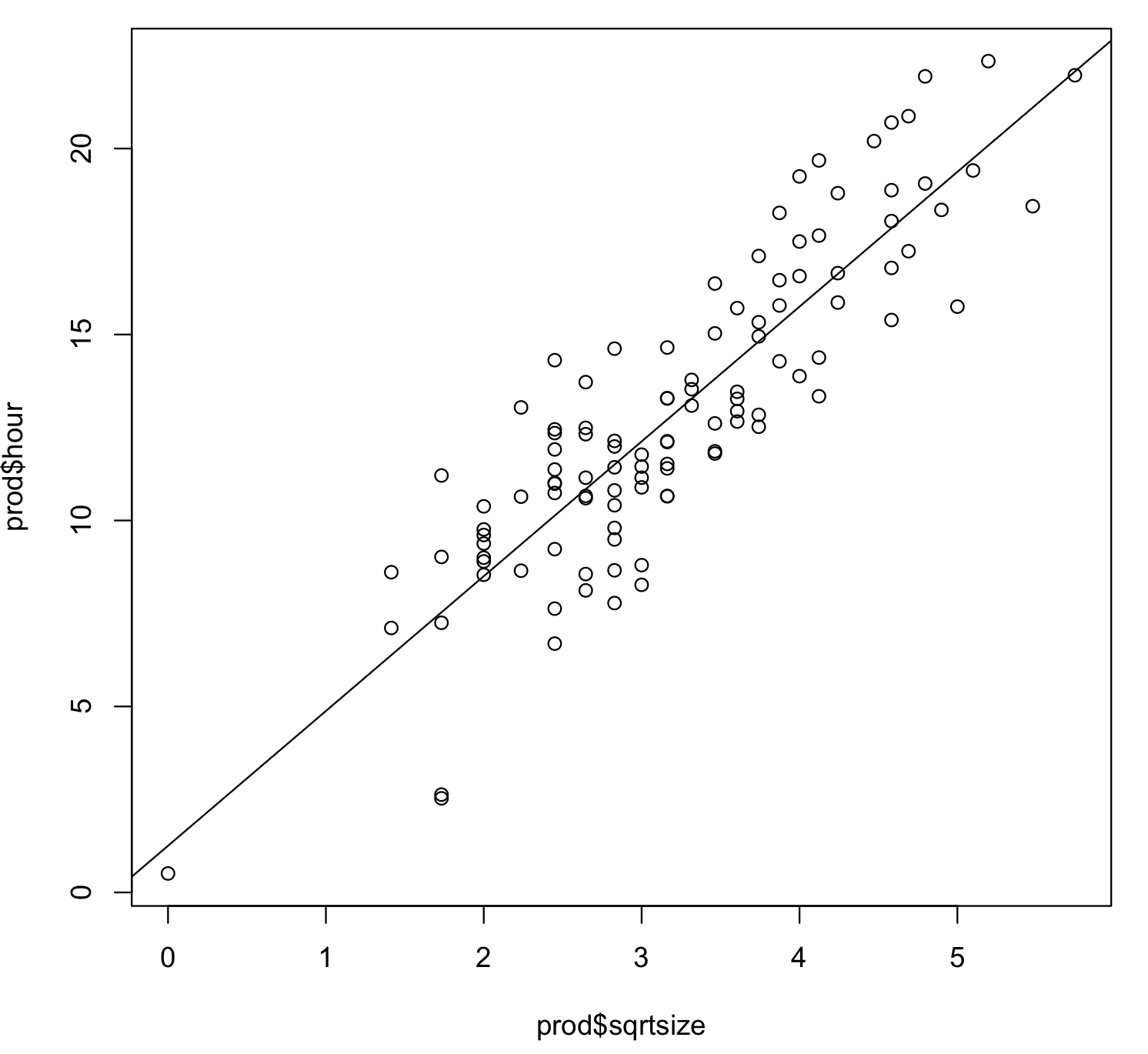
a.



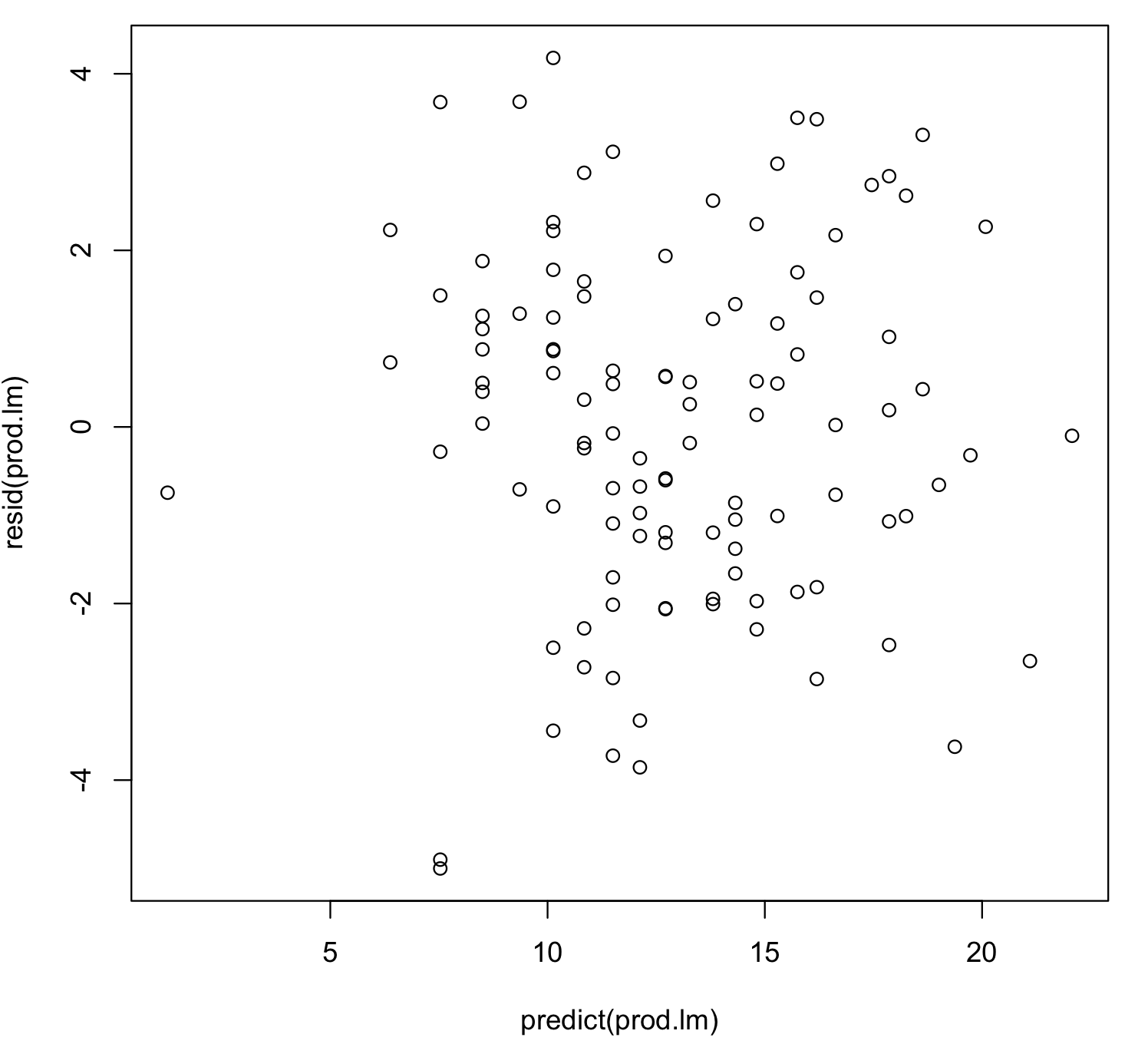
The relationship seems to be curvilinear, so a linear relation is not adequate here. We can try some transformation on X because the variability at different X level are fairly constant.

b. Ŷ = 1.2547 + 3.6235\*X’

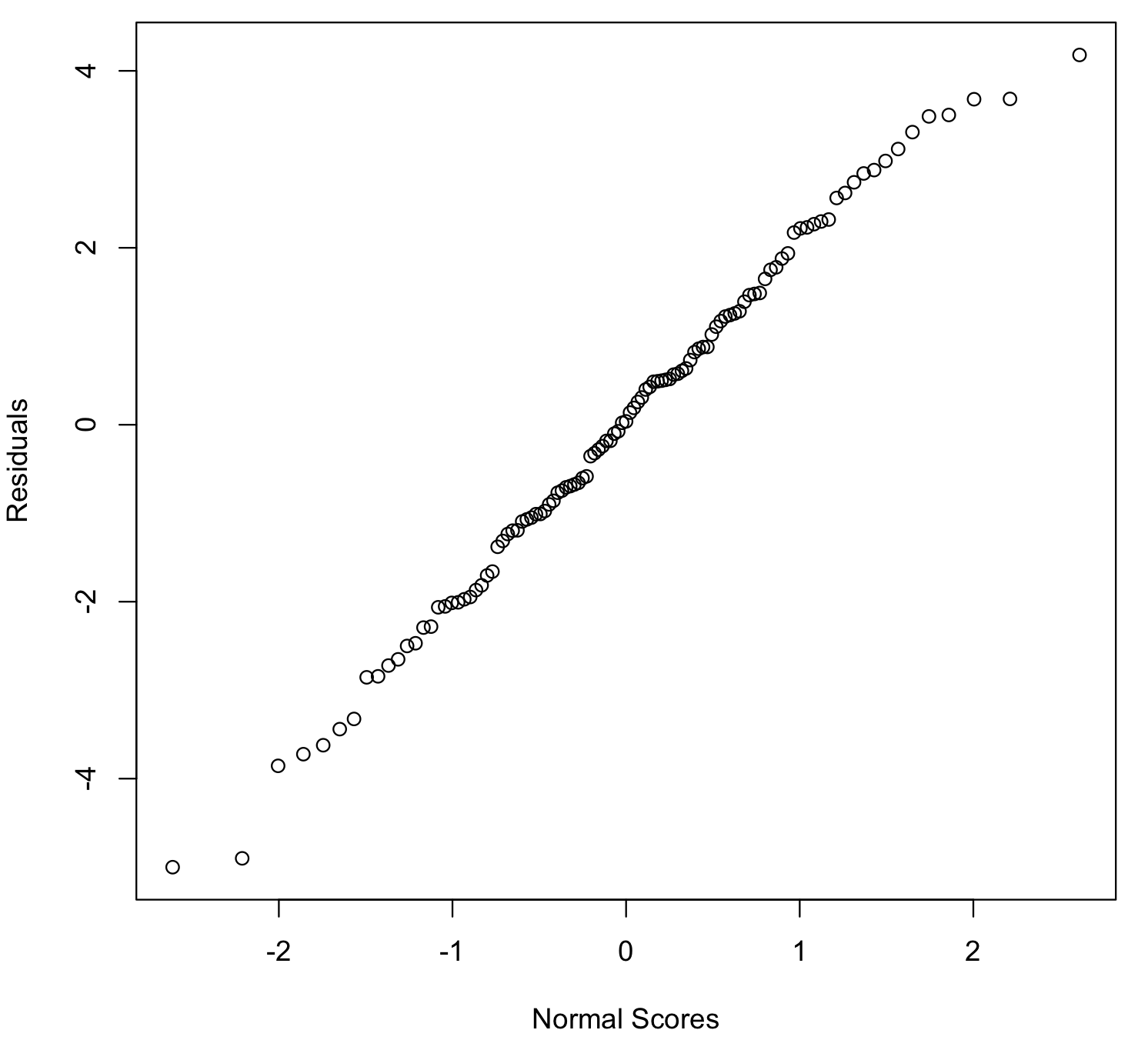
c. It appears to be a good fit.



d. Residuals against fitted values:



Normal probability plot:



The first plot shows that the residuals spread around 0 and has a fairly constant variance at different X level, and the second plot shows that the residuals approximately follow a normal distribution. These plots suggest that our transformation is appropriate.

e. Ŷ = 1.2547 + 3.6235 

Since our data contains lot of size 0, from the intercept of the estimated regression function, we estimate the production time of lot of size 0 is 1.2547 hours. The slope shows that, in our estimate, the production time increases by 3.6235 as the square root of the lot size increases by 1.

3.20

For X’ = 1/X: since this is just a linear transformation on X, the scale of Y does not change. Therefore, it does not change the variability of the error terms and the error terms will still be normally distributed.

For Y’ = 1/Y: this transformation might change the distribution of Y, so the variability of the error terms might change and the error terms might not be normally distributed.