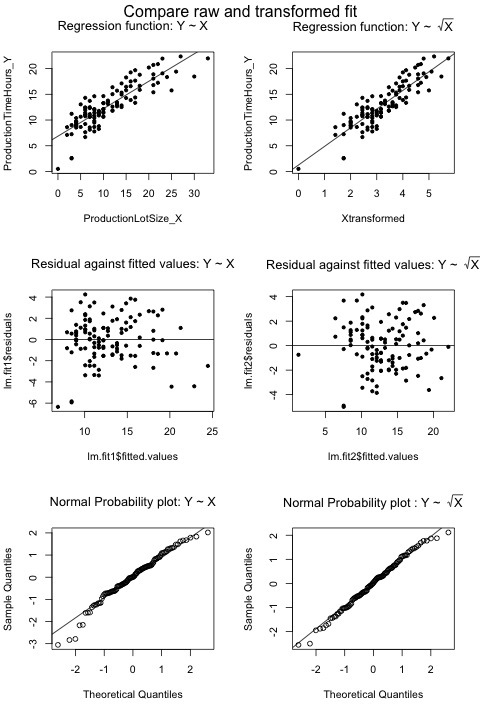
**Question 1. Textbook 3.18**



a) See the scatter plot above.

The simple linear regression fit is a little problematic, as the normal probability plot indicates that the data might not be normally distributed.

Transformation on X might be more appropriate, as the residual plot indicates that the variability is roughly constant, and transformation on Y might have negative influence on this pattern.

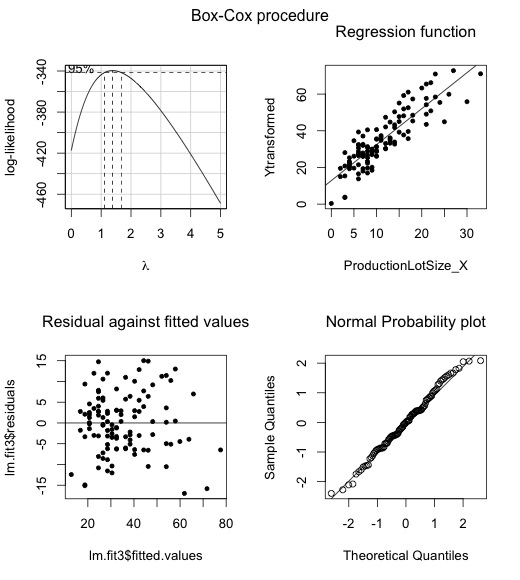
b) See the transformed regression plot above.

Estimated regression function:

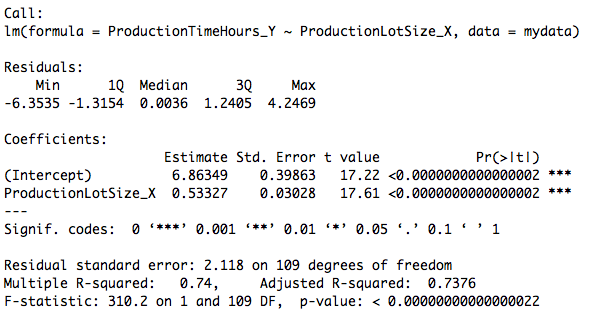
c) After transformation, the regression function fit the data pretty well, as there is no obvious pattern in the residual plot. Additionally, on the normal probability plot, the actual values become closer to the expected values under normality.

d) When we are only comparing the two residual plots, it is not obvious which regression model is better. Based on the normal probability plots, the transformation on X seems to provide a better fit with expected values. In particular, the normal probability plot for the raw data seems to deviate from the theoretical values around the lower left tail.

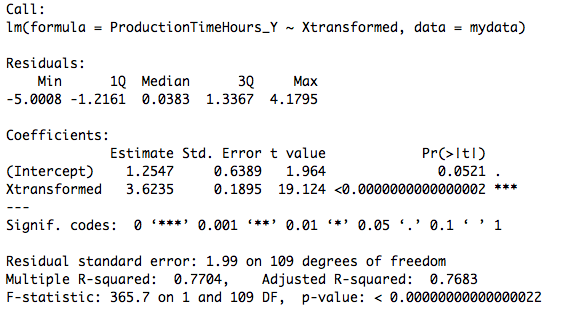
Use **Box-Cox procedure** to find transformation parameter for Y



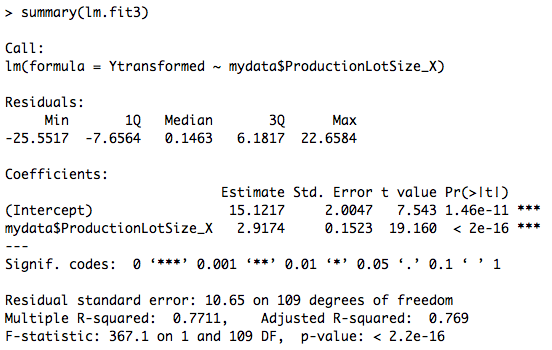
Summary for regression without transformation



Summary for regression with transformation:



Summary for regression with Box-Cox procedure:



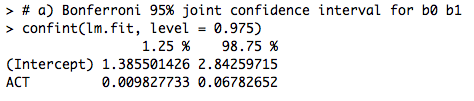
Conclusion:

Based on all the plots, it is not obvious whether Box-Cox procedure provided a good fit. However, the statistical results indicate that Box-Cox might unreasonable. Since it’s residual standard error (10.65) is the largest among all three models.

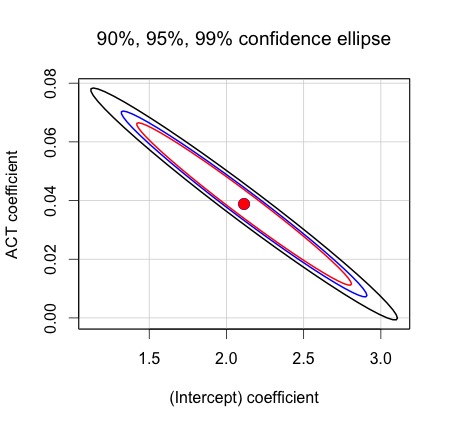
In general, the square root transformation seems to be the best.

**Question 2. GPA data**

1)



2) From inside to outside, the ellipses represent 90%, 95%, 99% joint confidence interval respectively.



3) When 0 is 2.0, 1 is unlikely to be 0.025, since this point is outside of 90% confidence ellipses.

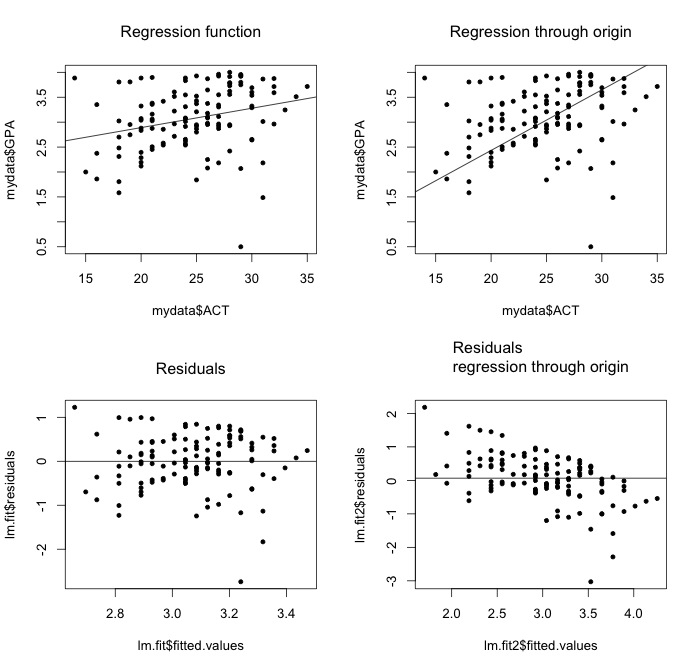
**Question 3, textbook 4.1**

When we obtained the 90% joint confidence intervals for 0 and 1 with Bonferroni method, it doesn’t mean that there is 10 percent chance for 0 to be incorrect, or there are 5 percent chance for 0 and 5 percent chance for 1 to be wrong.

If b0 and b1 has 5% chance to be wrong and these two evens are independent, then the probability of both b0 and b1 are correct would be:

In addition, this is because b0 and b1 are not independent so their covariance is non-zero, the joint confidence interval has a elliptical shape. If the covariance if zero, the ellipse will become a circle, then we might be able to argue that they are equally likely to be wrong with certain probability.

**Question 4, textbook 4.1**



1. When comparing the SLR with regression from the origin on the raw data, it is inconclusive whether the regression from the origin is a good fit here.
2. Based on the residual plot, we can be pretty sure that that regression through origin is a bad fit, as the residuals is not constant. In particular, the residuals are decreasing as the fitted values become larger.

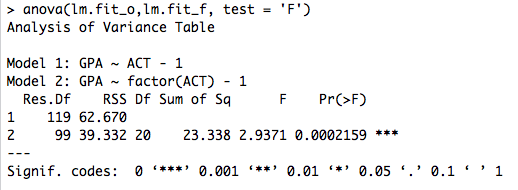
Based on computation, the residuals do not sum to zero.

Macintosh HD:Users:Qihong:Desktop:Screen Shot 2015-03-01 at 10.05.42 AM Mar 1.png

1. Lack of fit test for regression through origin model, a = 0.005

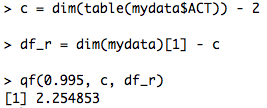
H0: E(Y) = 1 \* X

H1: E(Y) != 1 \* X



There are 21 different ACT scores so the df\_C = 21 – 2 = 19

There are 120 students so the df\_r = 120 – c = 101



In conclusion, there is statistical evidence indicates the regression function through origin lack of fit, *F*(19, 101) = 2.25483, *p* < 0.005.