STAT 2131
Orly Olbum

Homework 3

1. KNNL 3.14

Y = hardness of items
X = elapsed time since termination

(a) Perform F-test to determine if there's a lack of fit of linear regression function.

$H_0: E(Y) = \beta_0 + \beta_1 X$ (Y is expected to be linearly represented)
$H_A: E(Y) \neq \beta_0 + \beta_1 X$ (Y is NOT ↑)
For $\alpha = .01$, reject H_0 if $|F^*| > F(.99; 2, 12) = 6.927$

Find SSLF (lack of fit)
$SSLF = \sum_i \sum_j (\hat{Y}_{ij}^* - \bar{Y}_i^*)^2$ with $df = c-2$
and SSPE (pure error)
$SSPE = \sum_i \sum_j (Y_{ij}^* - \bar{Y}_i^*)^2$ with $df = n-c$
} computed with R software.

To get an F-statistic of:
$F^* = \frac{SSLF/c-2}{SSPE/n-c}$

Here, $SSLF = 17.675$, $SSPE = 128.750$
so $F^* = \frac{(17.675/12)}{(128.75/12)} = .8237$
Because $F^* < F(.99; 2, 12)$, we fail to reject the H_0 and conclude that we don't have sufficient evidence, at 99% confidence, that Y is NOT a good fit for linear regression.

(b) Because having replications of X leads to estimates of the means being the same precision level, there is no real advantage or disadvantage to having equal number of replications at each X level

→

1, continued

© If H_0 in @ had been rejected, the conclusion would not have indicated what type of regression equation would be appropriate, apart from it not being linear. Next steps would be to test other models and interpret further conclusions about H_0's of other models.

2. KNNL 3.18

$n = 111$

Y = time in hours

X = production lot size

@ (see next page for scatterplot)

A linear relationship does not appear to represent the relationship between production lot size and time. We may want to transform X and regress

ⓑ Transform $X \to \sqrt{X}$, obtain estimated linear regression equation

(using R)

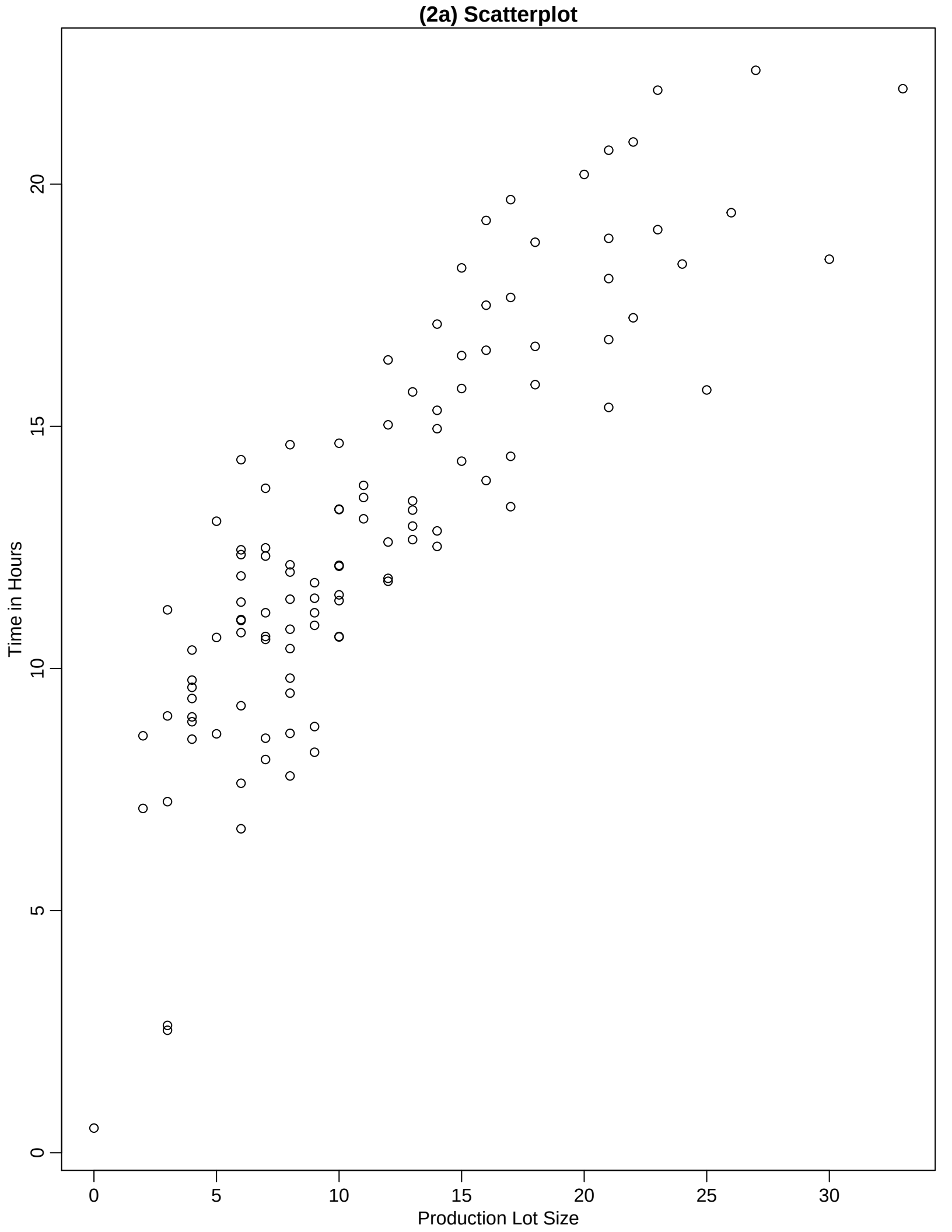
- create new variable Xt (X transformed)
- linear model

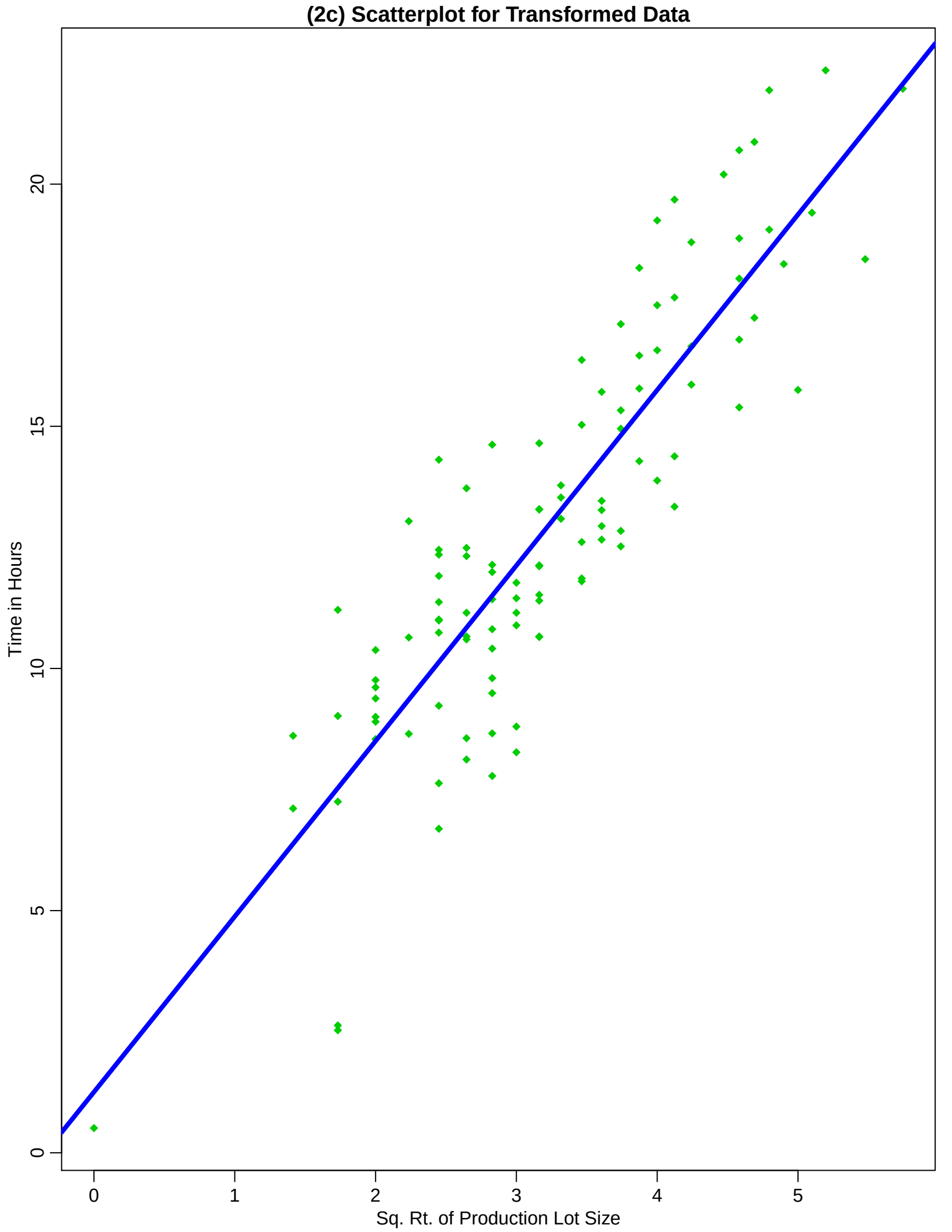
$$\hat{Y} = 1.2547 + 3.6235 X_t$$

© Plot transformed data & estimated regression line. (see next page for scatterplot)

The linear regression line does appear to be a good fit for the transformed data

→


(2a) Scatterplot
Time in Hours
Production Lot Size
0
5
10
15
20
25
30


(2c) Scatterplot for Transformed Data
Time in Hours
0
5
10
15
20
0
1
2
3
4
5
Sq. Rt. of Production Lot Size

2, continued

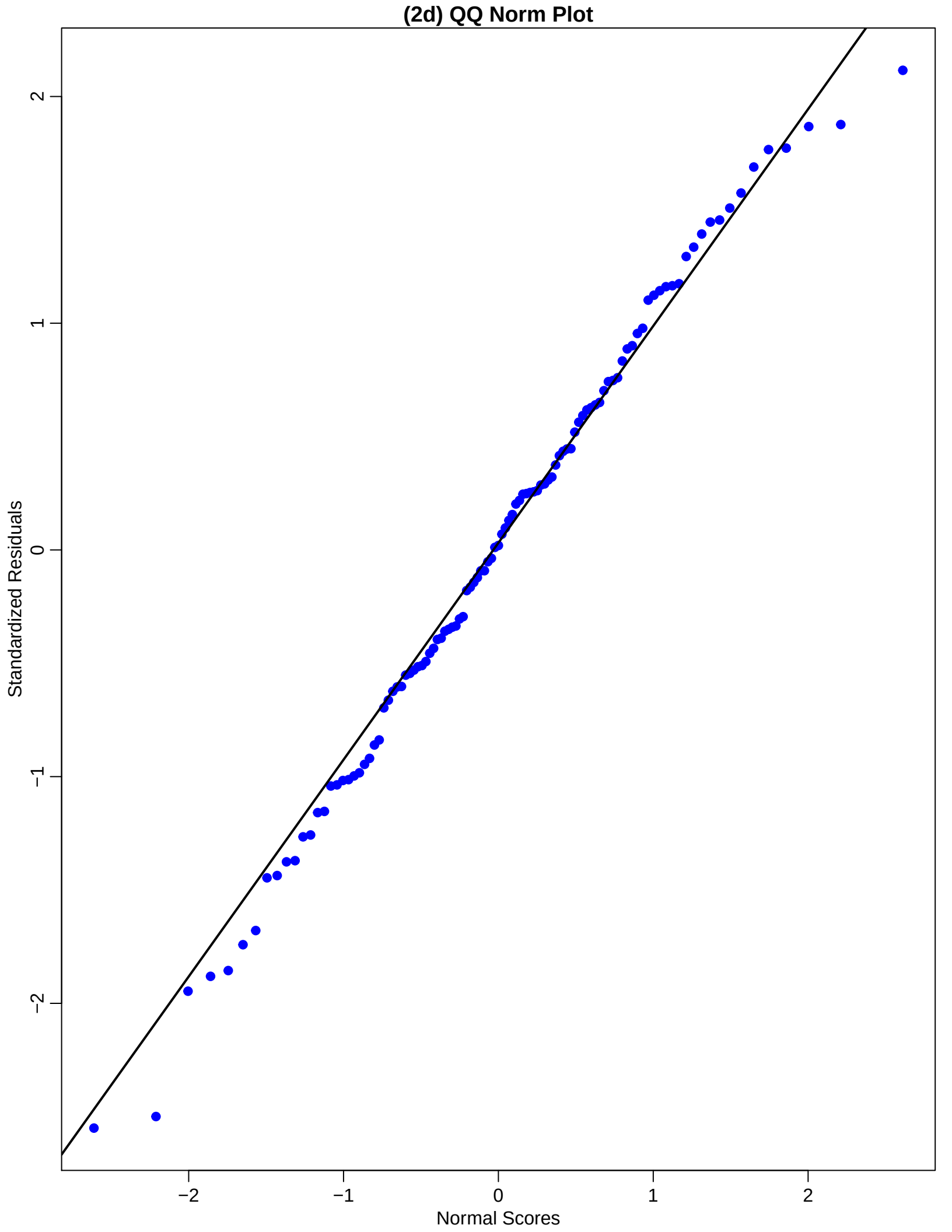
(d) Plot residuals & against fitted values, with QQ plot

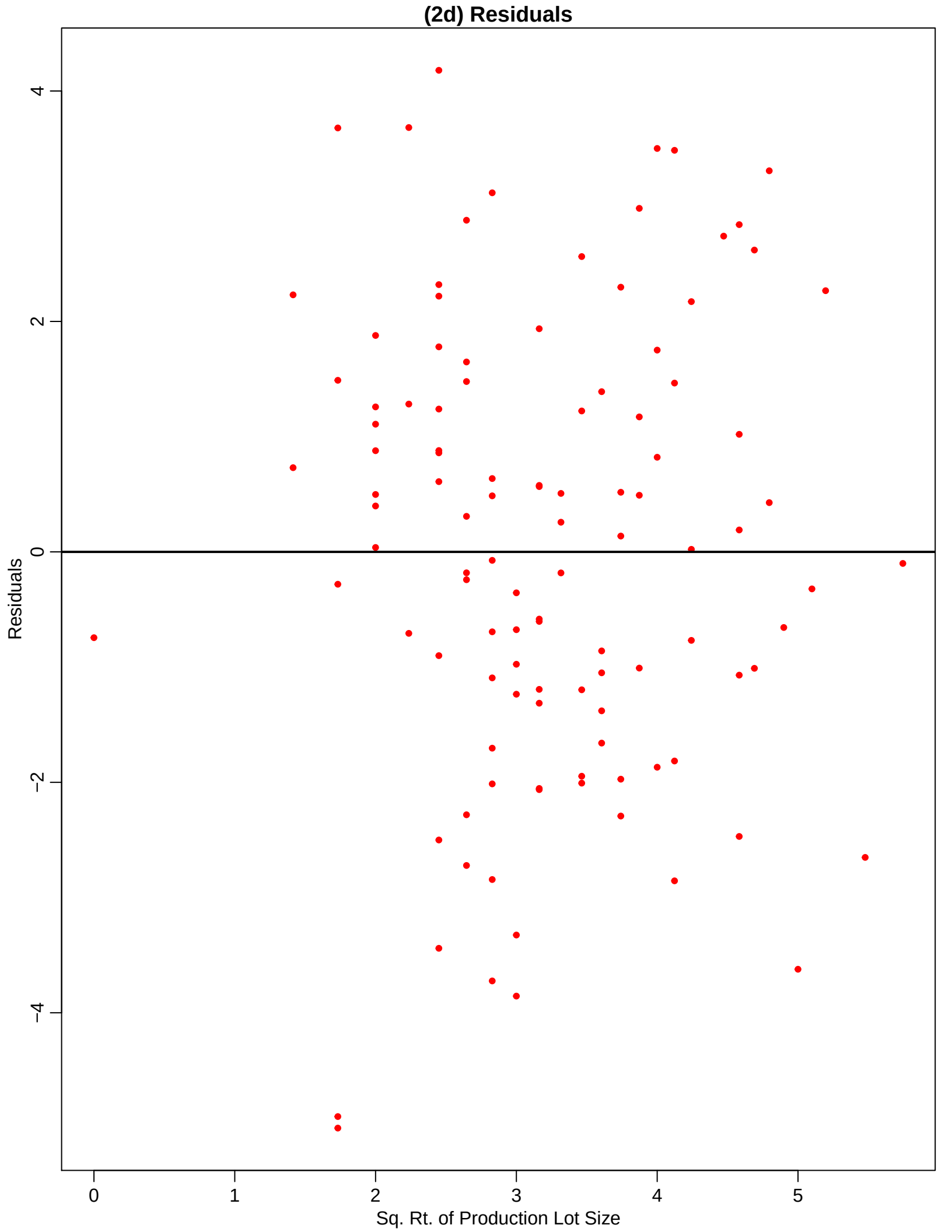
(see next page for plots and residuals)

The plots show scattered (read: uncorrelated!) residuals and qqplot almost directly on the qq trendline. This tells us that our variables come from the same distribution, indicating our [transformed] data is normally distributed

(e) Regression Function, untransformed

$$\hat{Y} = 1.2547 + 3.6235\sqrt{X}$$


(2d) QQ Norm Plot
Standardized Residuals
Normal Scores
2
1
0
−1
−2
−2
−1
0
1
2


(2d) Residuals
Residuals
4
2
0
−2
−4
0
1
2
3
4
5
Sq. Rt. of Production Lot Size