

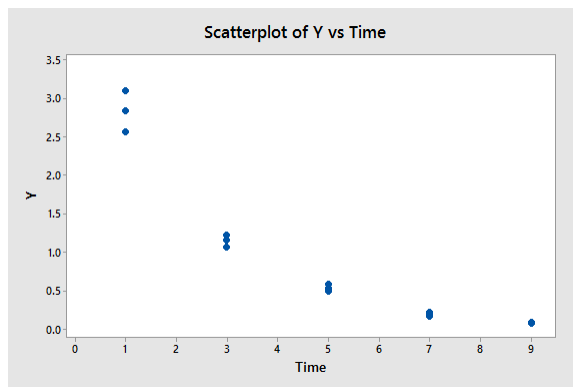
Homework 8

Roly Vicaría
STAT501 Fall 2015

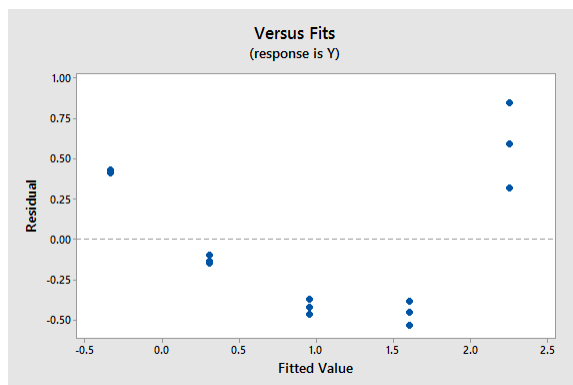
October 25, 2015

0.0.1 Question 1

- a) The plot shows a negative association between y and time. As time increases, the y decreases. The plot also shows a curvilinear relationship between y and time. There is also greater variance in the y values for smaller values of time.



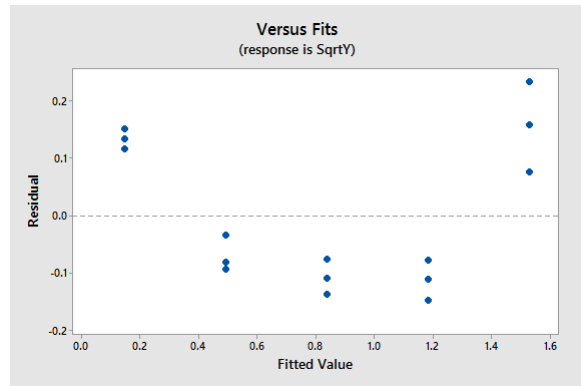
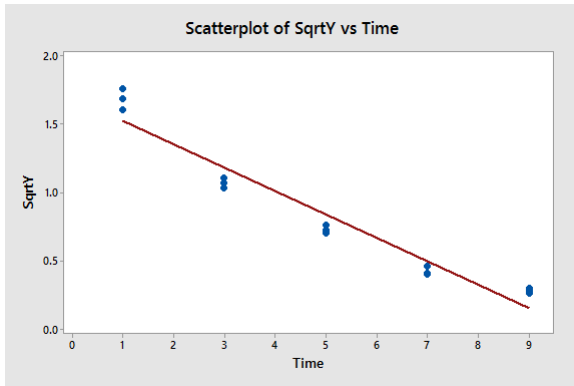
- b) $Y = 2.575 - 0.3240 \text{ Time}$



The plot of residuals vs fitted values indicates a non-linear relationship and non-constant variance of the errors.

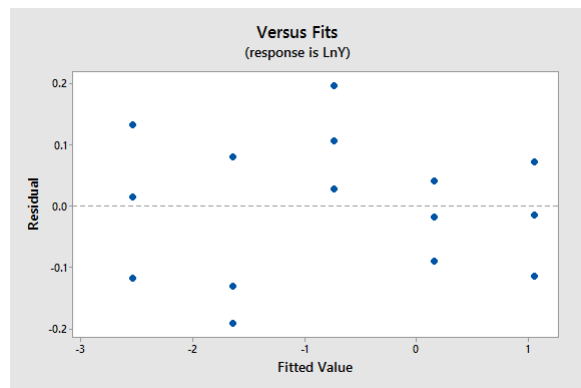
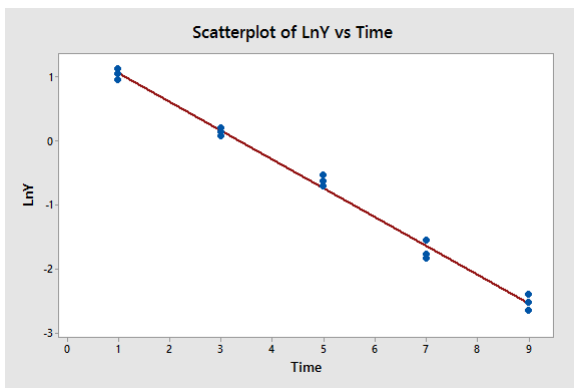
- c) Based on the two previous plots, it may be best to transform the y -variable in this case, because in this case there are issues with the linearity AND the error terms. A transformation on x primarily corrects the non-linearity, but does not do much to correct the variation of the error terms. A transformation on y corrects problems with the error terms and may help the non-linearity.

d) $SqrtY = 1.6998 - 0.1724 Time$



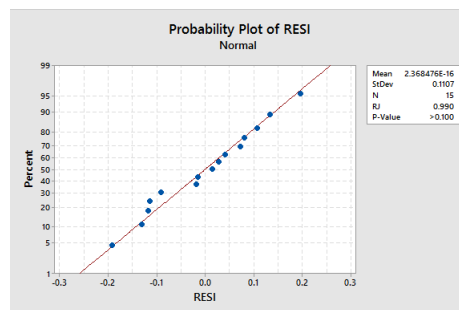
The plot of sqrtY vs time looks much better than the previous plot and shows an R-sq value of 93.85%. However, the plot of residuals versus fitted value still shows an up-down-up pattern indicating non-linearity.

e) $LnY = 1.5080 - 0.4500 Time$



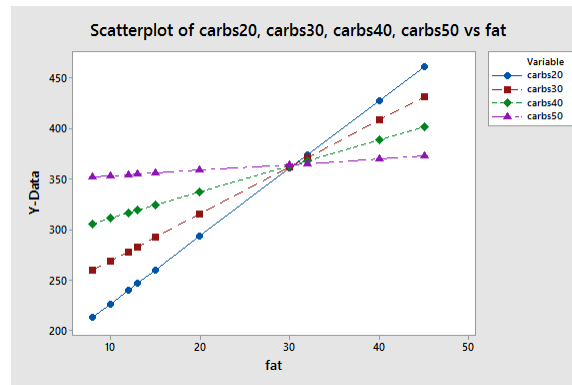
The plot of lnY vs time looks almost perfectly linear compared to the previous plots and shows an R-sq value of 99.30%. The plot of residuals versus fitted values now shows more of a horizontal band shape with points distributed pretty evenly above and below the residual = 0 line. This indicates good linear fit. The error variance appears constant and normally distributed.

- f) Of the models examined, the model fitting lnY versus time is the best. It's residual versus fit plot showed the most constant error variance and normal distribution. The Ryan-Joiner test shows a high p-value so we fail to reject the null hypothesis that the terms are normally distributed. The other models all showed signs of non-linearity and non-constant variance of the error terms.



0.0.2 Question 2

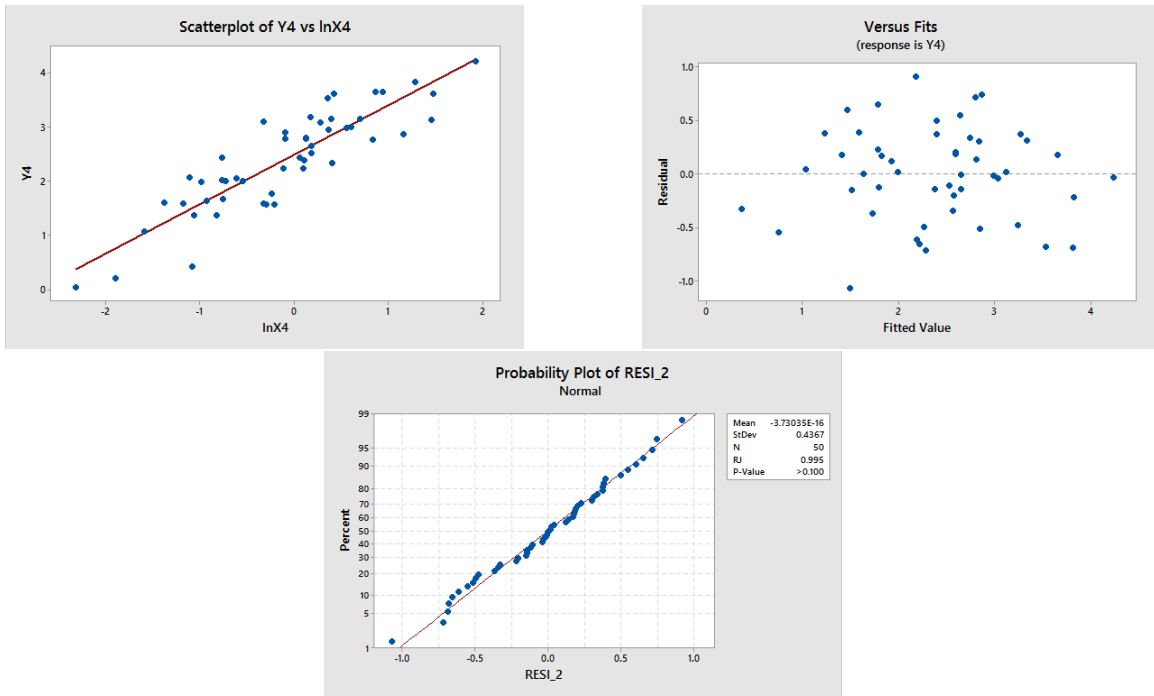
- a) $Calories = 33.6 + 6.28 carb + 10.81 fat - 0.2053 carb * fat$
- b) Minitab shows a p-value of 0.034 for the estimate of β_3 . At the 5% significance level, we reject the null hypothesis that $\beta_3 = 0$ and conclude that β_3 is significant. In the context of this problem, β_3 signifies the effect that carb and fat have on Calories in the presence of each other.
- c) i. $Calories = 159.2 + 6.704 fat$
 ii. $Calories = 222 + 4.651 fat$
 iii. $Calories = 284.8 + 2.598 fat$
 iv. $Calories = 347.6 + 0.545 fat$
- d) Plots of above equations for carbs = 20, 30, 40, and 50:



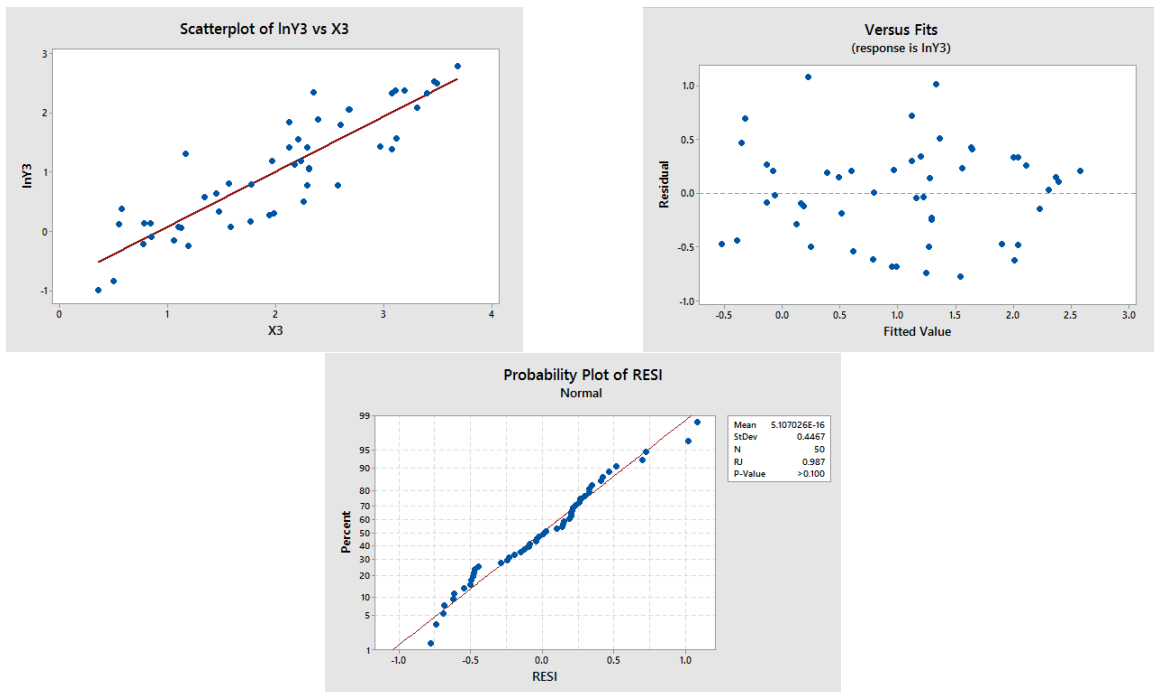
- e) The model equation estimated in part (a) shows us that the interaction effect between carbs and fat intake on the amount of calories generated is that as one of the values increases, the relative contribution of the other value decreases. In other words, the equations from part (c) showed that as we increased the value of carbs by units of 10, the slope on fat decreased by roughly 2.05. The lines showed that at higher carb values, calories hardly changed in response to fat. This relationship would be similar if we fixed fat and plotted out functions of carbs.

0.0.3 Question 3

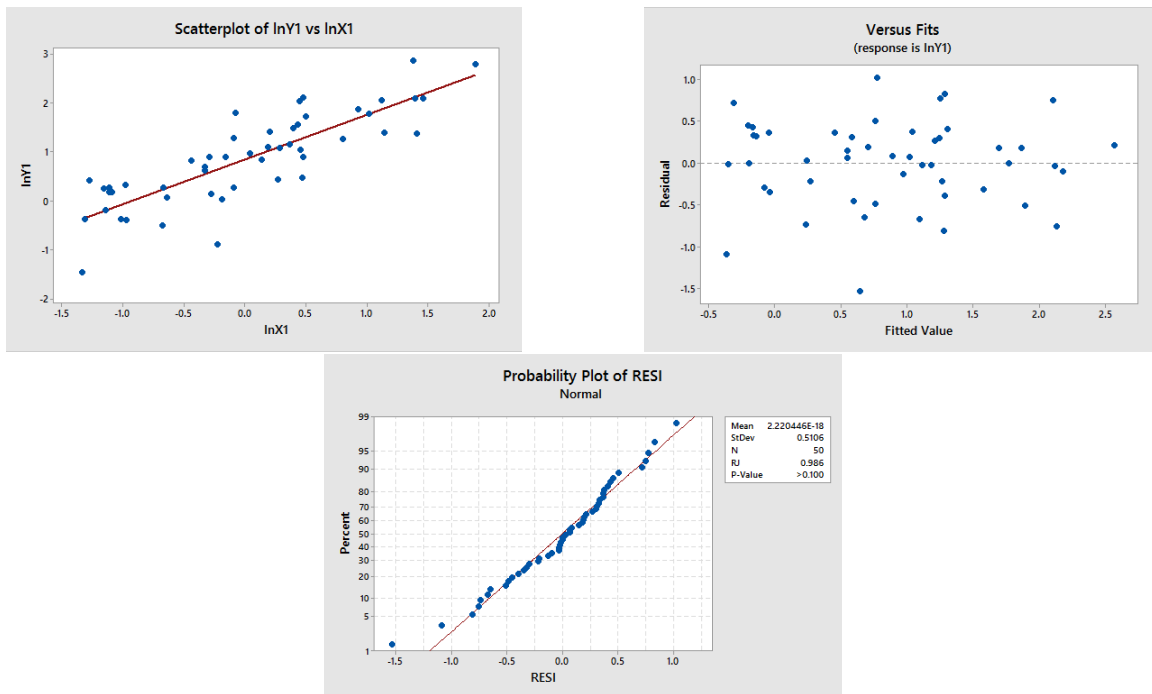
a) Response variable y , predictor variable $\ln(x)$: (y_4, x_4)



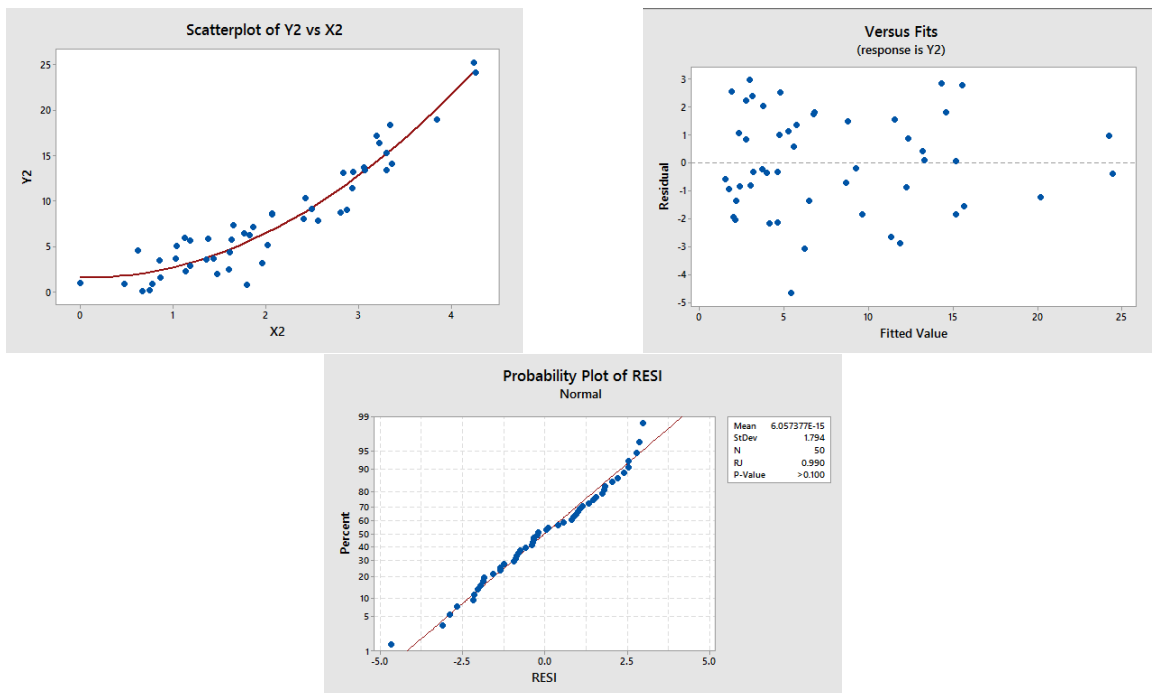
b) Response variable $\ln(y)$, predictor variable x : (y_3, x_3)



c) Response variable $\ln(y)$, predictor variable $\ln(x)$: (y_1, x_1)

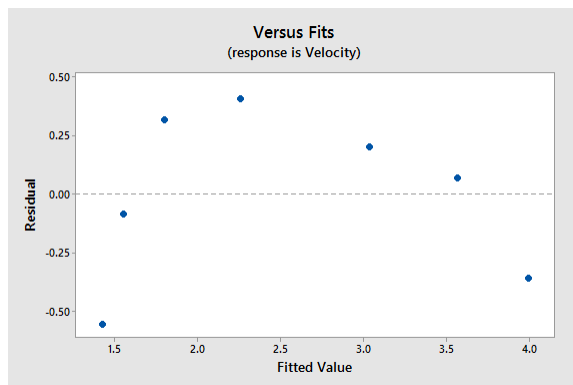


d) Response variable y , predictor variables x and x_2 : (y_2, x_2)

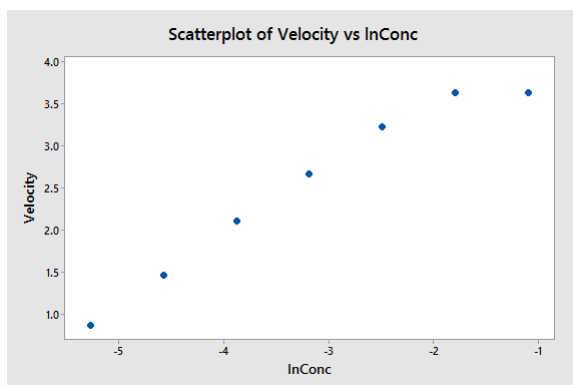


0.0.4 Question 4

- a) $Velocity = 1.288 + 25.65 \text{ Concentration} - 56.5 \text{ Concentration}^2$
- b) Test statistic: t-statistic = -3.33, d.f. = 5, p-value = 0.029. The p-value indicates that at significance level of less than .05, we can reject the null hypothesis that $\beta_2 = 0$ and conclude that $\beta_2 \neq 0$. In the context of this problem, it indicates that velocity has a significant relationship to the square of concentration.
- c) The plot of residuals vs fitted values for the quadratic model shows a down-up-down pattern which points to an obvious violation of the linearity assumption.



- d) $Velocity = 3.920 + 0.048 \ln Conc - 0.1041 \ln Conc * \ln Conc$



- e) Test statistic: t-statistic = -4.04, d.f. = 5, p-value = 0.016. The p-value indicates that at significance level of less than .05, we can reject the null hypothesis that $\beta_2 = 0$ and conclude that $\beta_2 \neq 0$. In the context of this problem, it indicates that velocity has a significant relationship to the square of the natural log of concentration.
- f) The 95% prediction interval for velocity at a concentration level of e^{-3} is (2.47741, 3.20441). This can be interpreted as saying that we can be 95% confident that the velocity of a chemical reaction involving a concentration level e^{-3} of substrate will be between 2.47741 and 3.20441.