Namrata Kolla
Homework 5
March 5, 2015

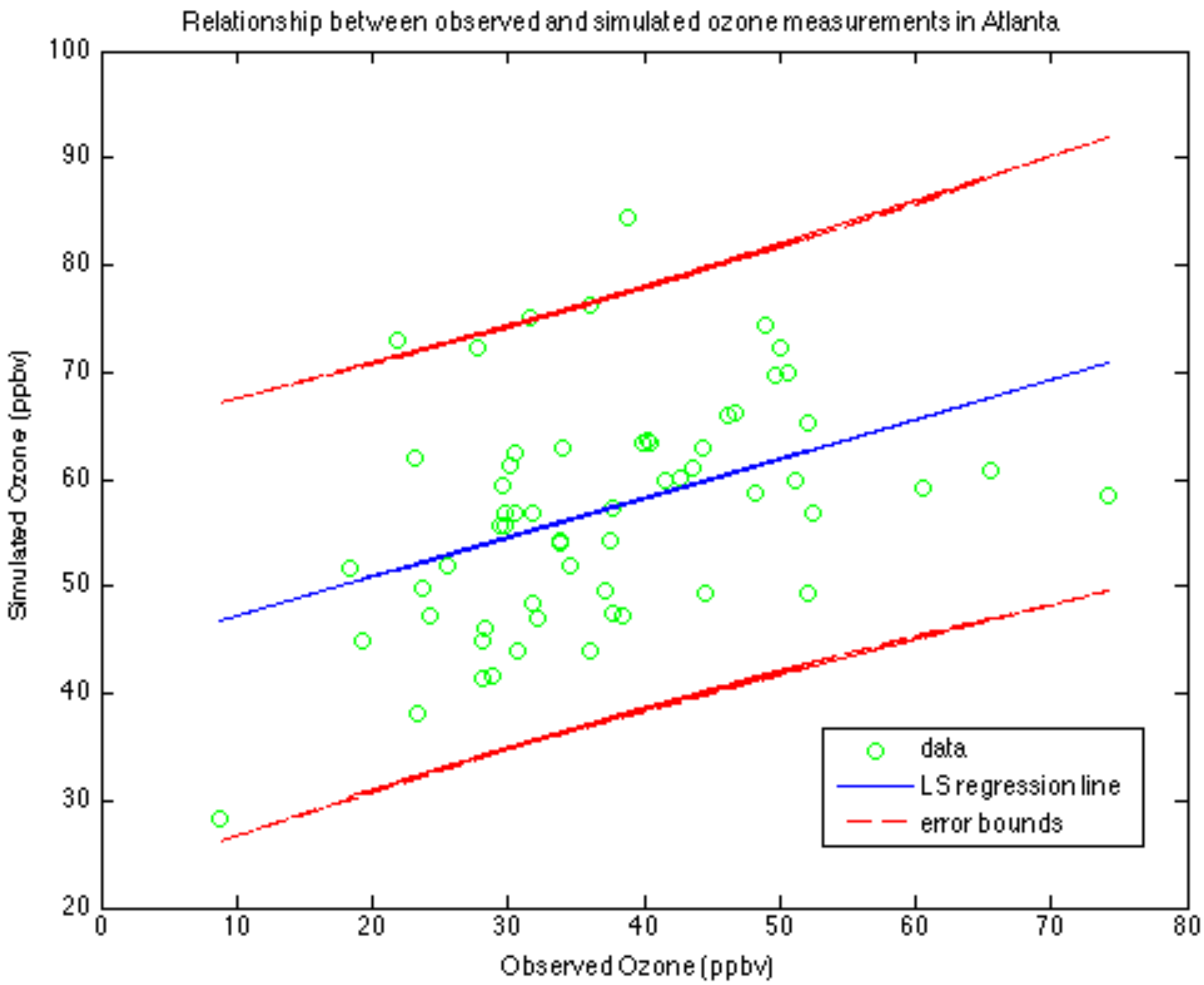
Problem 1

Part A: Least-squares regression and correlation coefficient

(1) Slope = 0.3665; y-intercept = 43.5618
95% confidence interval of the slope = [-1.2727, 2.0057]

(2) Correlation coefficient = 0.4114
95% confidence interval of correlation coefficient = [0.1759, 0.6024]
Correlation is significant because p = 0.0011 < 0.05

(3)



Comments on calculation methods:

To get the slope and y-intercept of the linear regression, I simply used the polyfit(function. To get the error bounds, I used the t-statistic by calculating error variance, calculating standard deviation for slope, multiplying those two values together, and adding/subtracting that value from the slope.

Part B: Through-the-origin least-squares regression

(1)

PART B

(1) $y_i = \beta_0 + \beta_1 x_i \longrightarrow y_i = \beta_1 x_i$

cost function $= Q = \sum (y_i - \beta_1 x_i)^2$

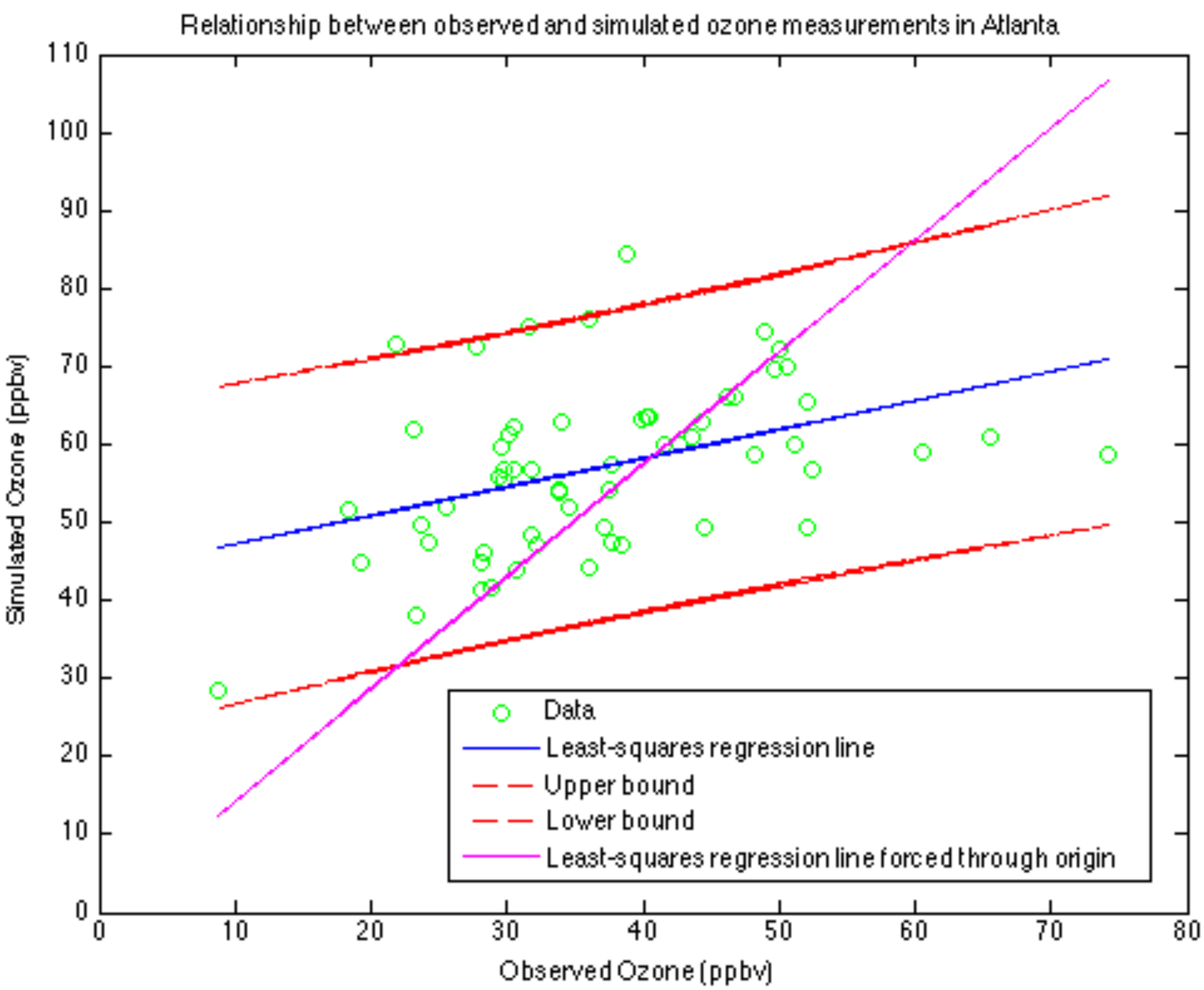
$dQ/d\beta_1 = -2\sum (y_i - \beta_1 x_i)(x_i)$

Setting $dQ/d\beta_1 = 0$: $\quad -2\sum (y_i - \beta_1 x_i)(x_i) = 0$

$\sum (y_i - \beta_1 x_i)(x_i) = 0 \quad \sum x_i y_i - \beta_1 x_i^2 = 0 \quad \sum x_i y_i = \sum \beta_1 x_i^2$

$\beta_1 = \dfrac{\sum x_i y_i}{\sum x_i^2} \quad y = \left(\sum \dfrac{x_i y_i}{x_i^2}\right) x$

(2)/(3)



(4)

(4) $(y_i - b) = \beta_0 + \beta_1(x_i - a)$

cost function $= Q = \Sigma[(y_i-b) - \beta_0 - \beta_1(x_i-a)]^2$

$dQ/d\beta_1 = 2\Sigma[(y_i-b)-\beta_0-\beta_1(x_i-a)][-(x_i-a)]$

$= -2\Sigma(x_i-a)[(y_i-b)-\beta_0-\beta_1(x_i-a)]$

setting $dQ/d\beta_1 = 0$:

$$\Sigma(x_i-a)(y_i-b) - \beta_0(x_i-a) - \beta_1(x_i-a)^2 = 0$$

$$\Sigma\beta_1(x_i-a)^2 = \Sigma(x_i-a)(y_i-b) - \beta_0(x_i-a)$$

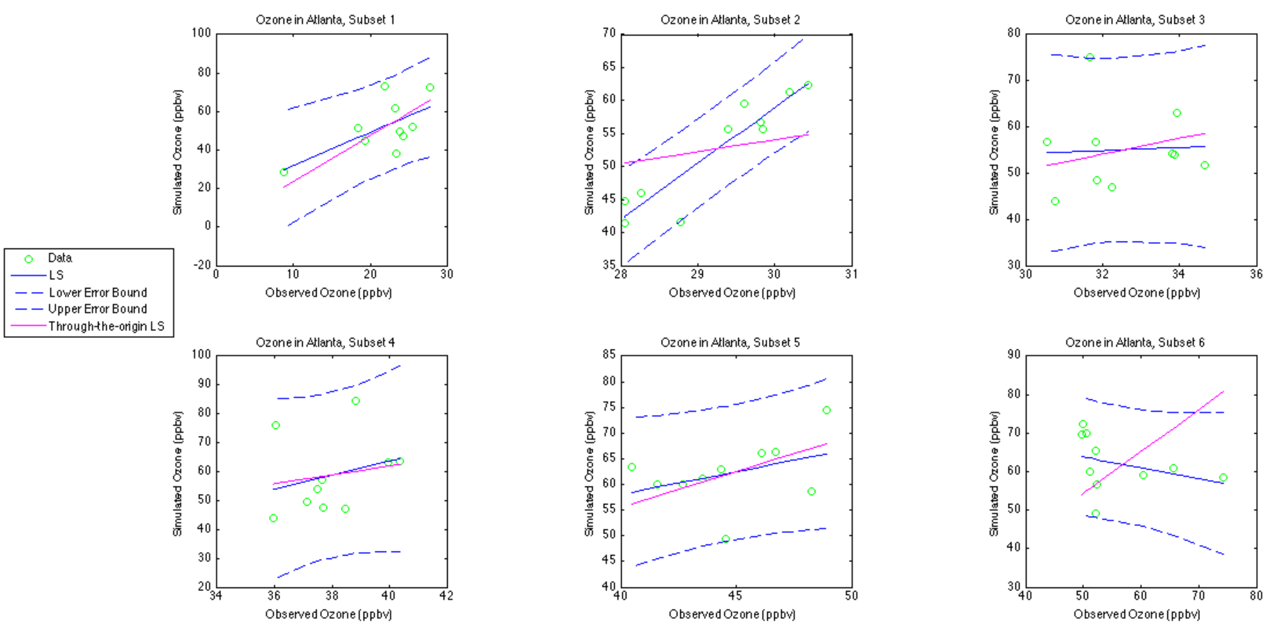
$$\beta_1 = \frac{\Sigma \cancel{(x_i-a)}(y_i-b) - \beta_0\cancel{(x_i-a)}}{\Sigma(x_i-a)^{\cancel{2}}} = \Sigma\frac{(y_i-b)-\beta_0}{(x_i-a)}$$

$$y = \left(\Sigma \frac{(y_i-b)-\beta_0}{(x_i-a)}\right)x$$

Part C: Resampled statistics

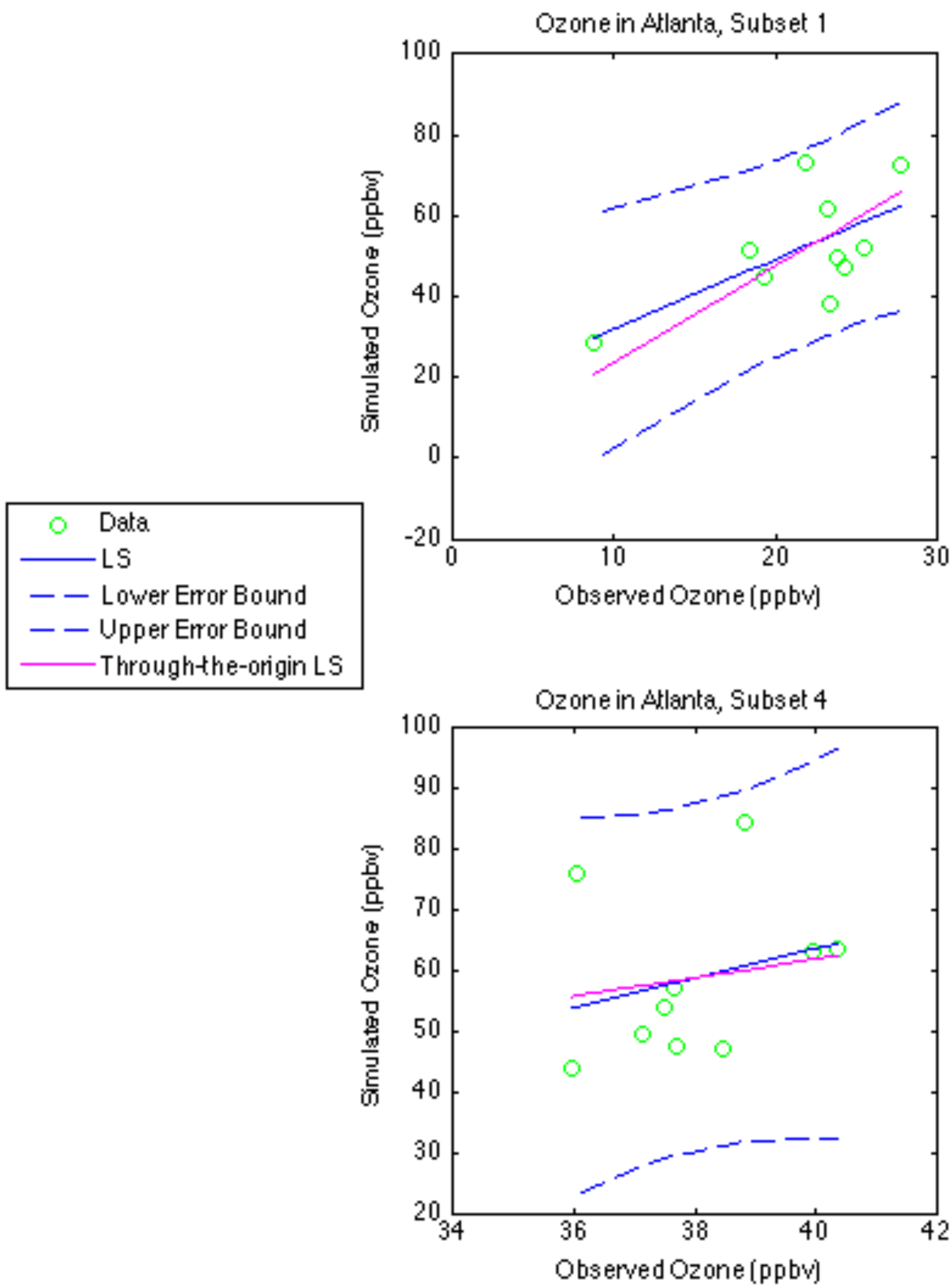
Subset	LS slope	LS confidence interval	Through-the-origin slope
1	1.6967	[-3.3206, 6.7139]	2.3670
2	8.5125	[0.2772, 16.7477]	1.8013
3	0.3539	[-14.7947, 15.5025]	1.6917
4	2.3957	[-19.1092, 23.9007]	1.5484
5	0.8829	[-4.2985, 6.0642]	1.3899
6	-0.2813	[-2.2710, 1.7084]	1.0868

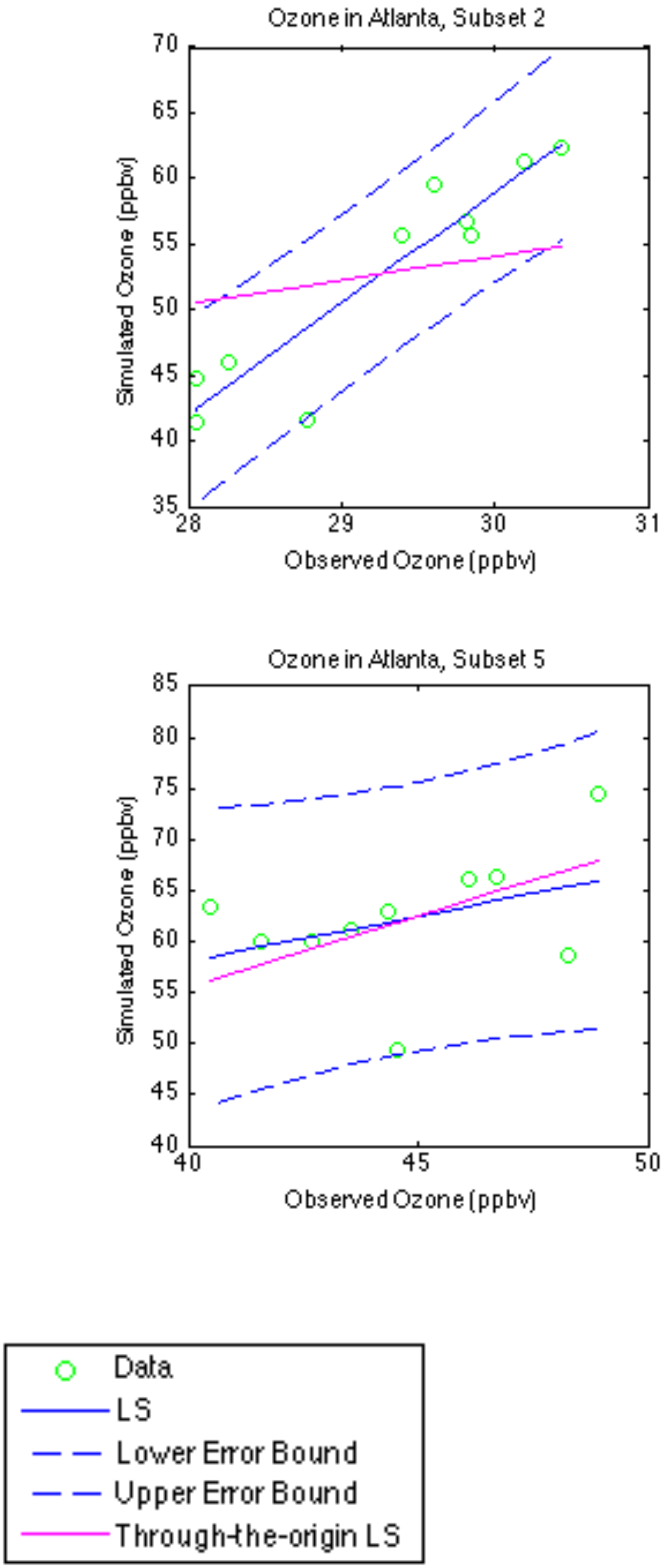
All plots:
(zoomed in versions in following pages)



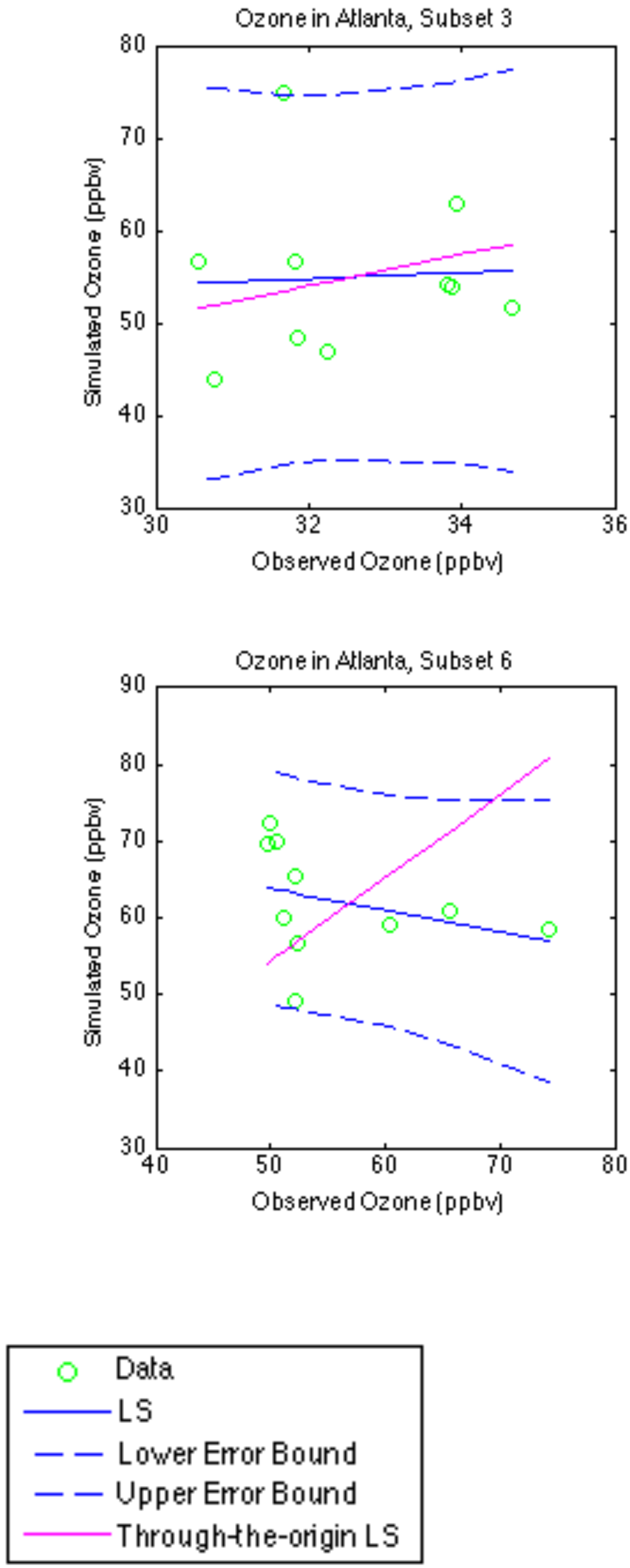
Comments on variation between subplots:

If the model worked well, then the slope would be near 1 because modeled values would be very close to measured values. When the slopes are not forced through the origin, subset 5's model has the best performance. Subset 6 was the only one with a negative regression slope, but it had the smallest confidence interval. This means that the model did a good job of mirroring the measured values, but it was consistently lower than the measured values.





Ozone in Atlanta, Subset 2
Simulated Ozone (ppbv)
Observed Ozone (ppbv)
70
65
60
55
50
45
40
35
28
29
30
31
Ozone in Atlanta, Subset 5
Simulated Ozone (ppbv)
Observed Ozone (ppbv)
85
80
75
70
65
60
55
50
45
40
40
45
50
Data
LS
Lower Error Bound
Upper Error Bound
Through-the-origin LS


Ozone in Atlanta, Subset 3
Simulated Ozone (ppbv)
Observed Ozone (ppbv)
30
40
50
60
70
80
30
32
34
36
Ozone in Atlanta, Subset 6
Simulated Ozone (ppbv)
Observed Ozone (ppbv)
30
40
50
60
70
80
90
40
50
60
70
80
Data
LS
Lower Error Bound
Upper Error Bound
Through-the-origin LS

Problem 2
(1)/(2)

PROBLEM 2

(1) $\sum_{i=1}^{m} \left(\frac{w - x_i}{\sigma_i}\right)^2 = Q$ where Q = cost function

$\frac{dQ}{dw} = 2\sum_{i=1}^{m} \left(\frac{w - x_i}{\sigma_i}\right)\left(\frac{1}{\sigma_i}\right) = 2\sum_{i=1}^{m} \left(\frac{w - x_i}{\sigma_i^2}\right)$ minimizing cost function: $\frac{dQ}{dw} = 0$

$\sum_{i=1}^{m} \left(\frac{w - x_i}{\sigma_i^2}\right) = \sum_{i=1}^{m} \frac{w}{\sigma_i^2} - \frac{x_i}{\sigma_i^2} = 0$ $\sum_{i=1}^{m} \frac{w}{\sigma_i^2} = \sum_{i=1}^{m} \frac{x_i}{\sigma_i^2}$

$w = wam = \frac{\sum_{i=1}^{m} x_i/\sigma_i^2}{\sum_{i=1}^{m} 1/\sigma_i^2}$

(2) variance $= \sigma^2$ $\sum_{i=1}^{m} w\sigma_i^2 = \frac{\sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} 1} = \sum_{i=1}^{m} x_i$ $\sigma_i^2 = \frac{\sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} w}$

(3) weighted average mean (wam) = 45.8878
uncertainty of wam = variance = 2.6704