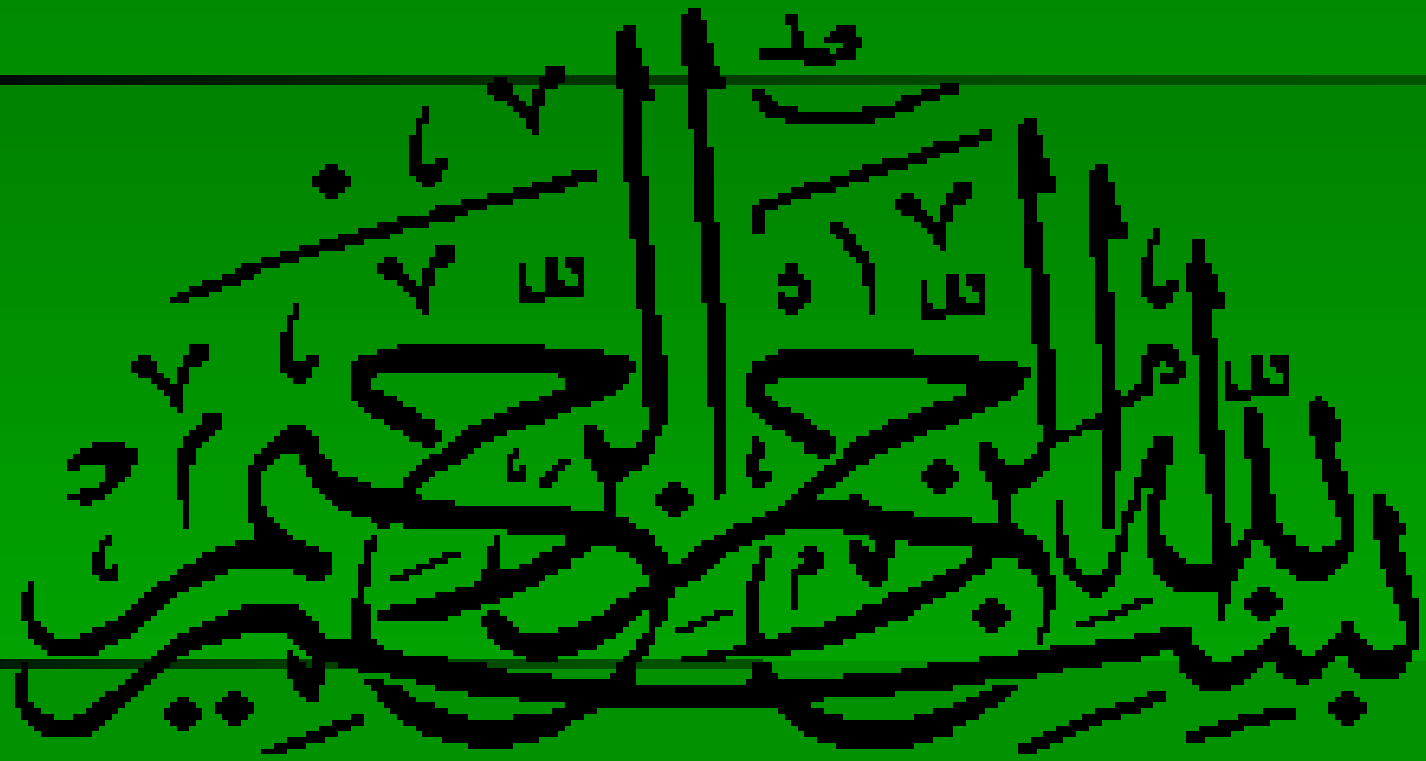


Applied Econometrics

Fall 2014, PIDE: Dr Asad Zaman

Lecture 5:

Looking at Data

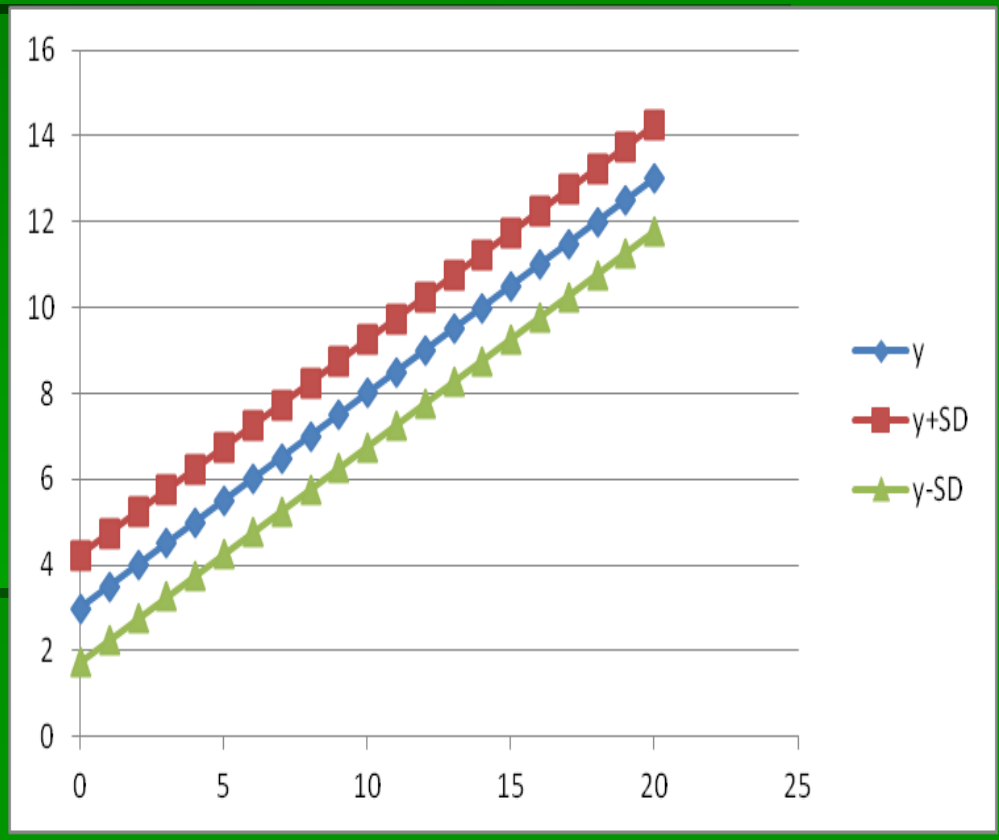
Data is All Important

Regression:

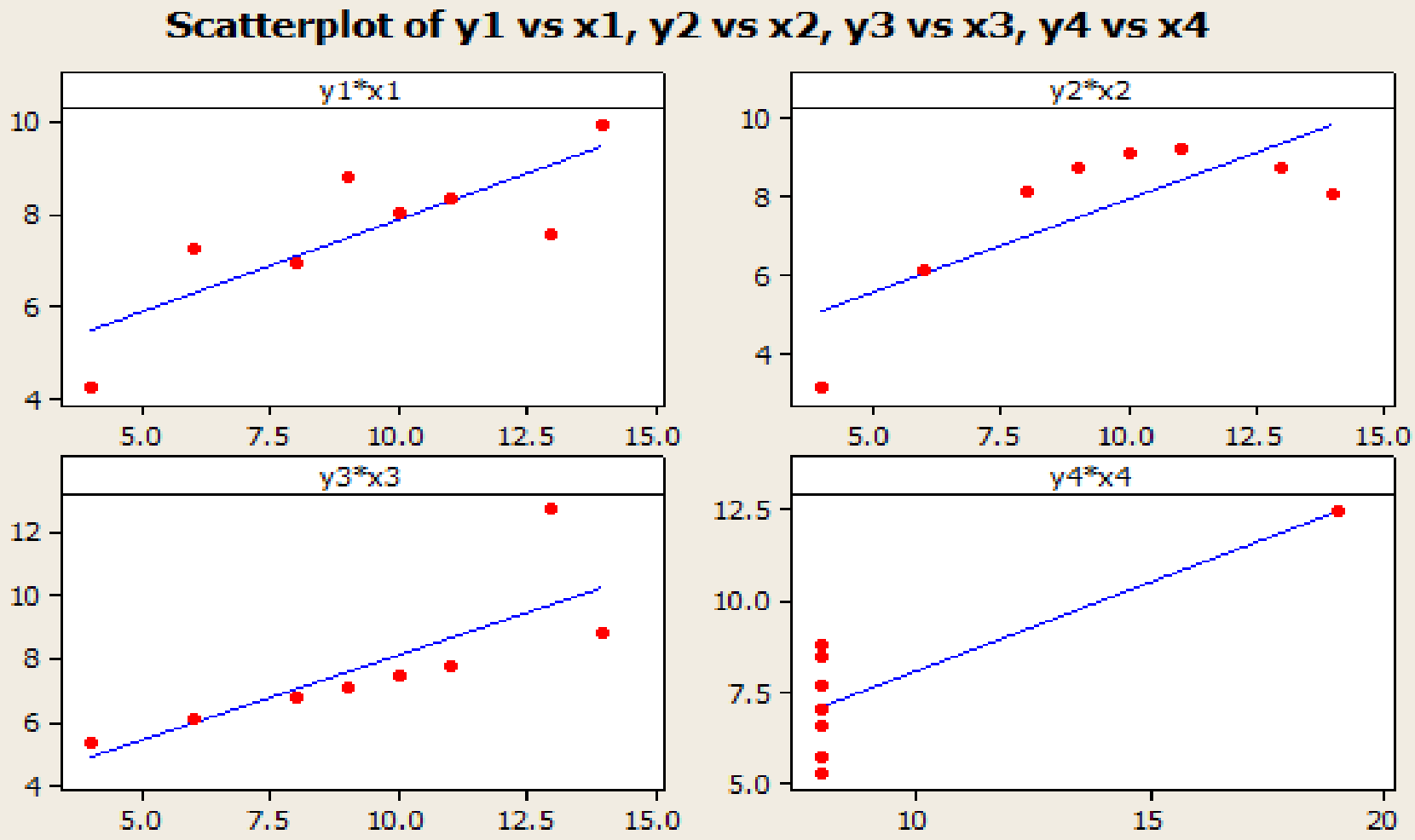
The regression equation

y1 = 3.00 + 0.500 x1

- SE = 1.23660
- R-Sq = 66.7%
- R-Sq(adj) = 62.9%



All 4 Datasets have Same Regression



Regression is WRONG Description of Data

- ~~Perfect fit to a quadratic.~~
- Perfect fit to a line with an outlier.
- Zero Fit + outlier.
- First picture: potential match for model.

Lesson: Regressions Meaningless without EDA

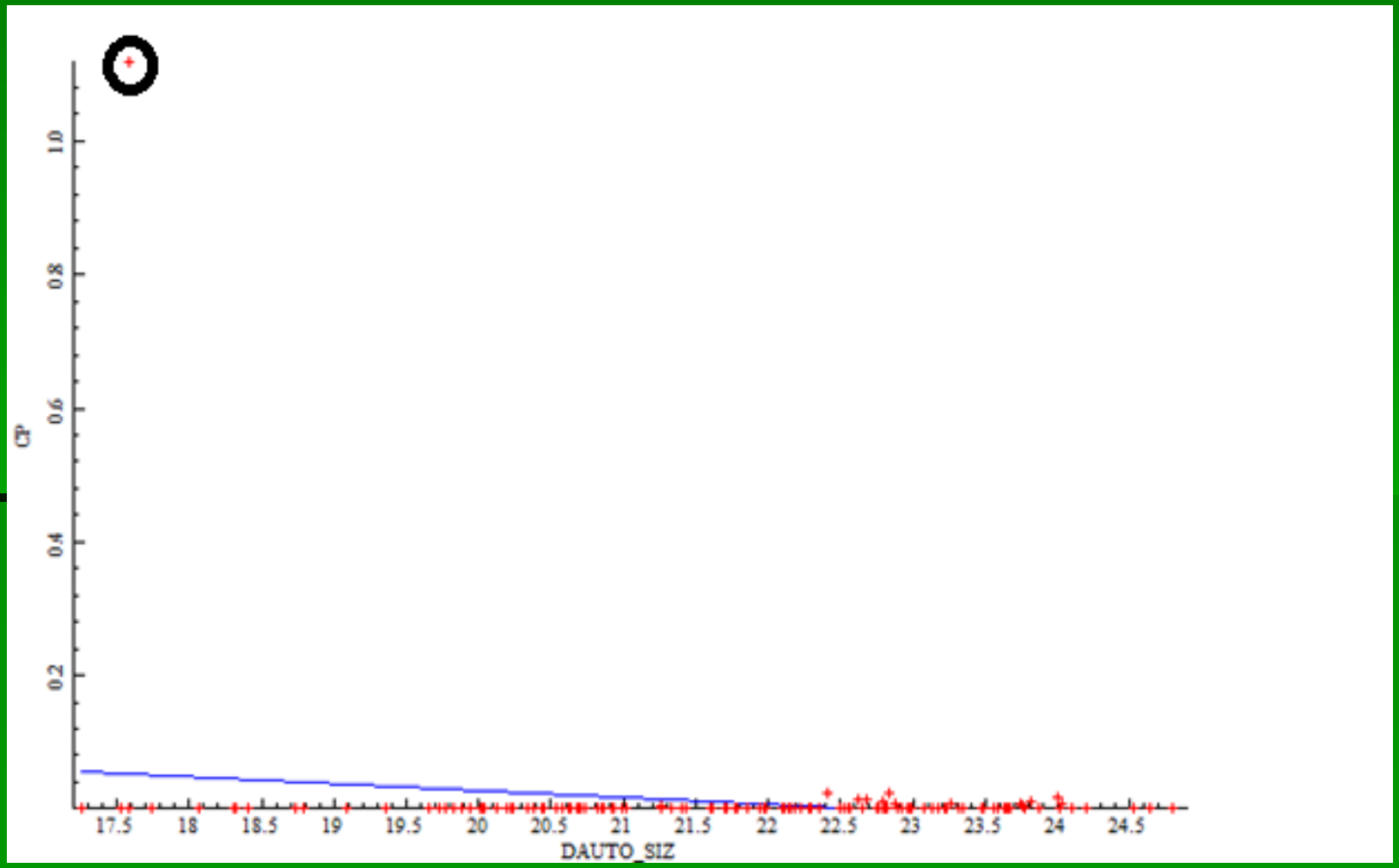
~~How to CHECK for~~

- Functional Form
- Outliers
- Clusters

========== ALL are types of:

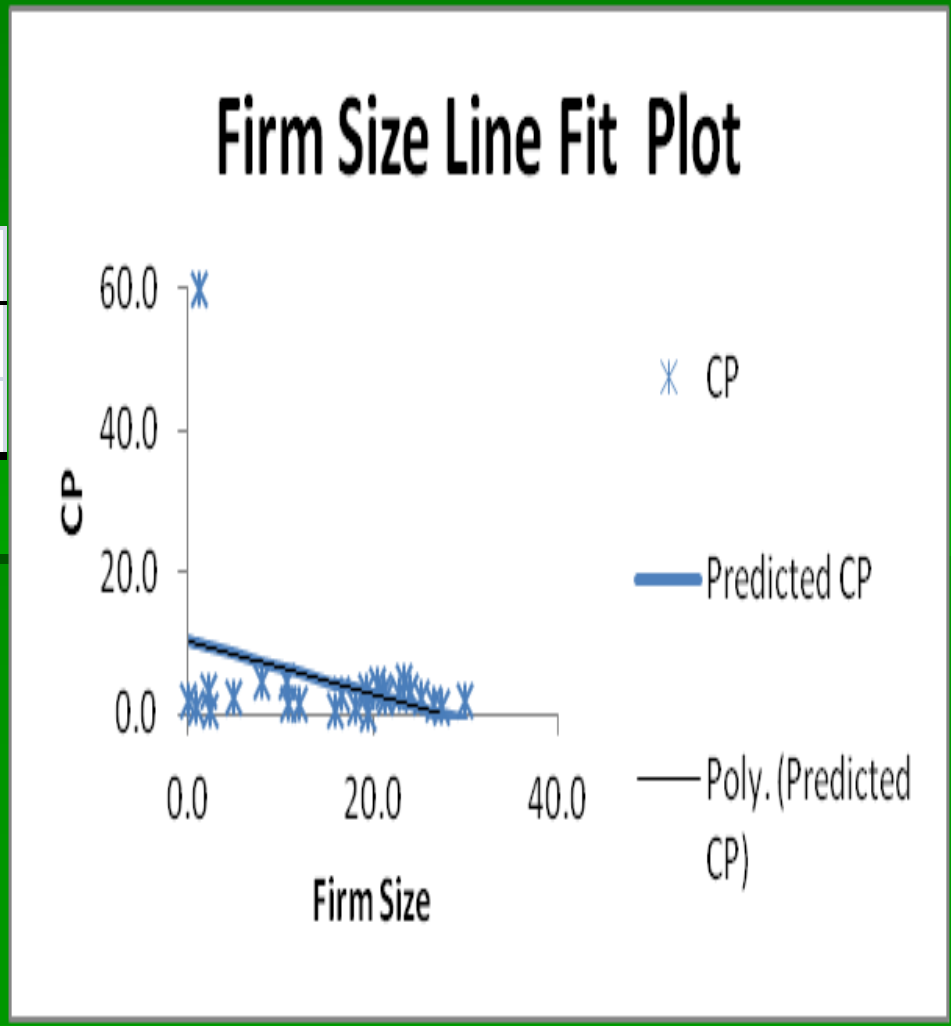
- MISSING EXPLANATORY VARIABLES.

Firm size and CP: Automobile Sector

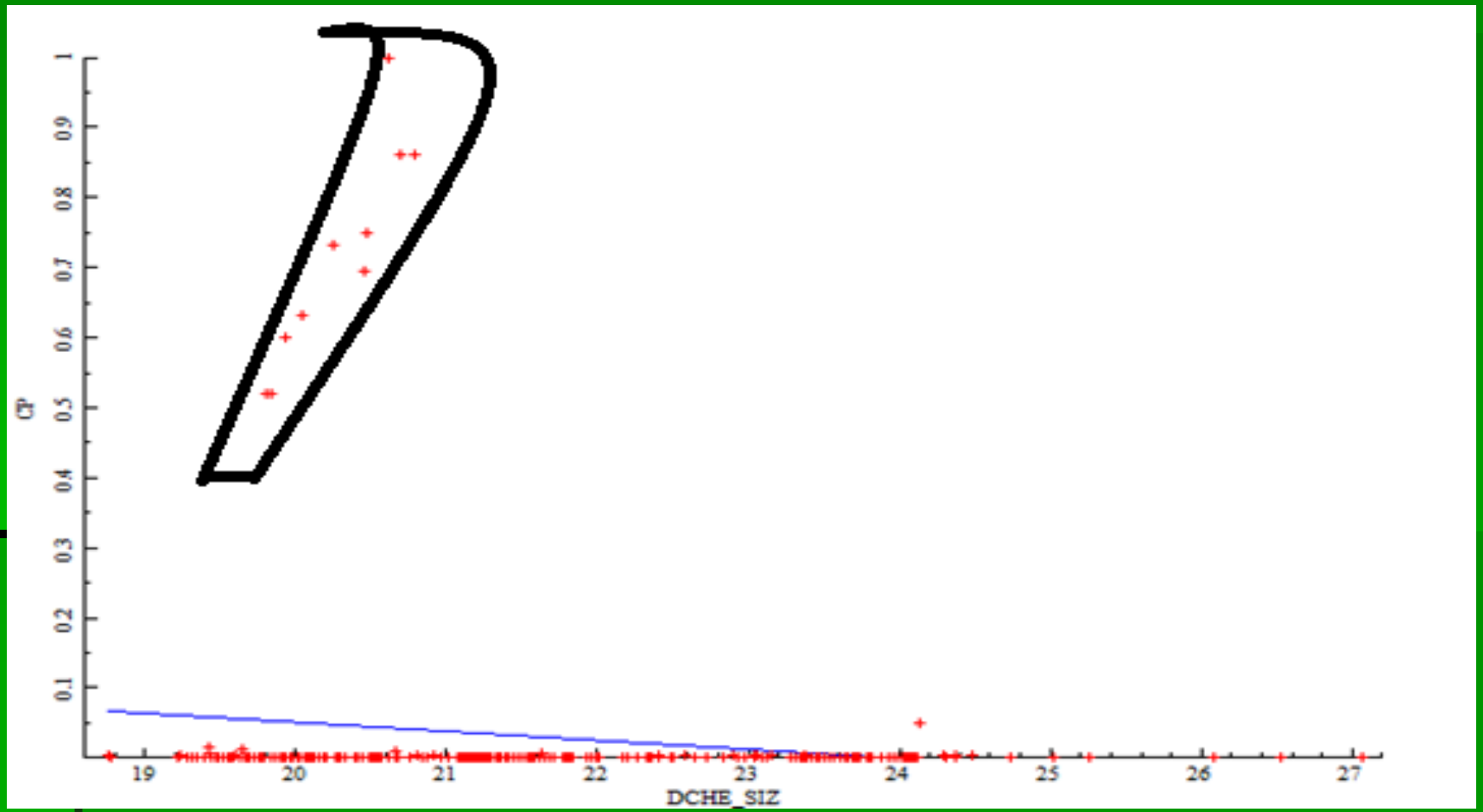


Simulated Data - Outliers

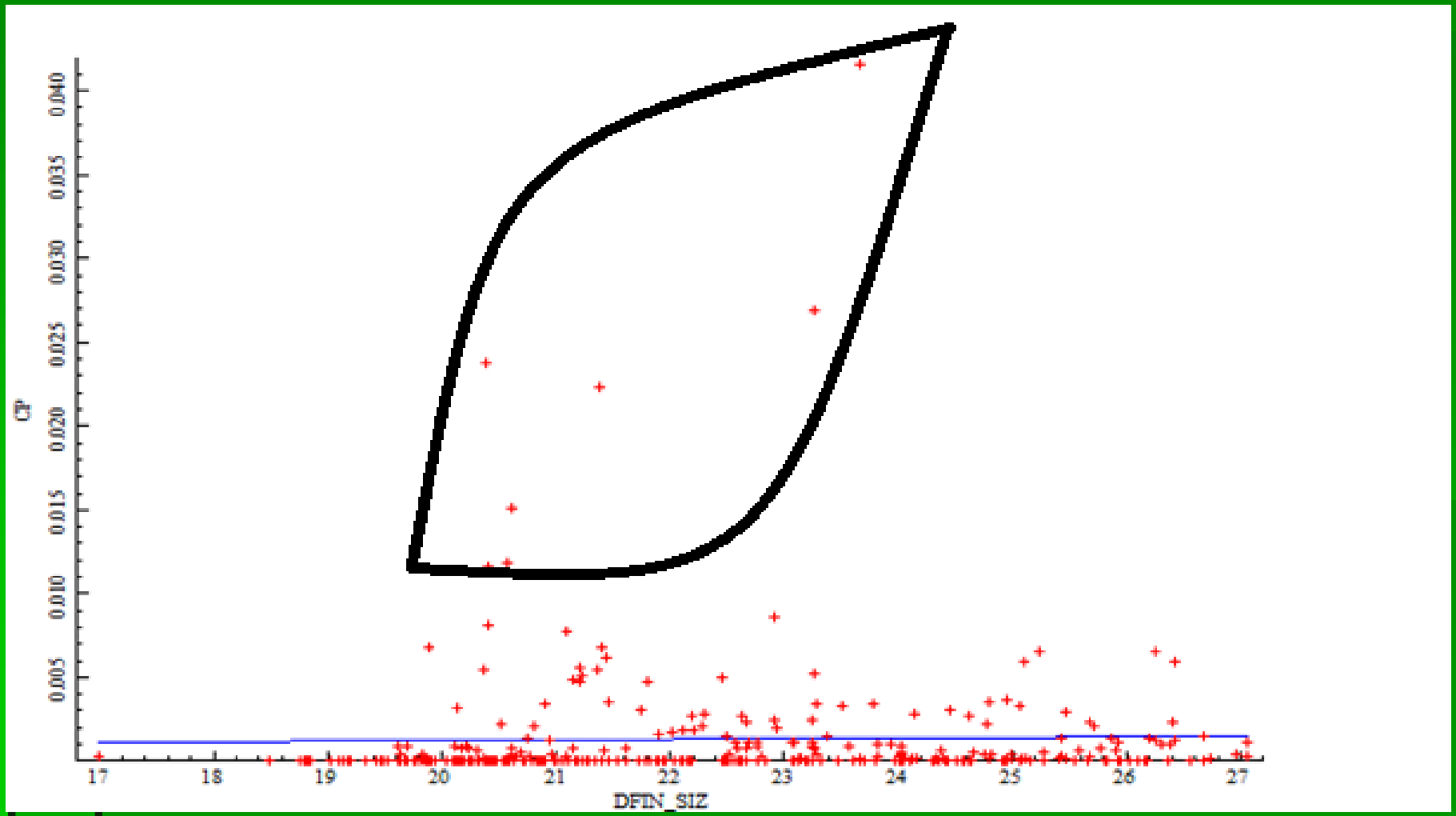
	Coefficients	*Standard Err*	*t Stat*
Intercept	10.23254	4.413563	2.318431
Firm Size	-0.37483	0.245897	-1.52434



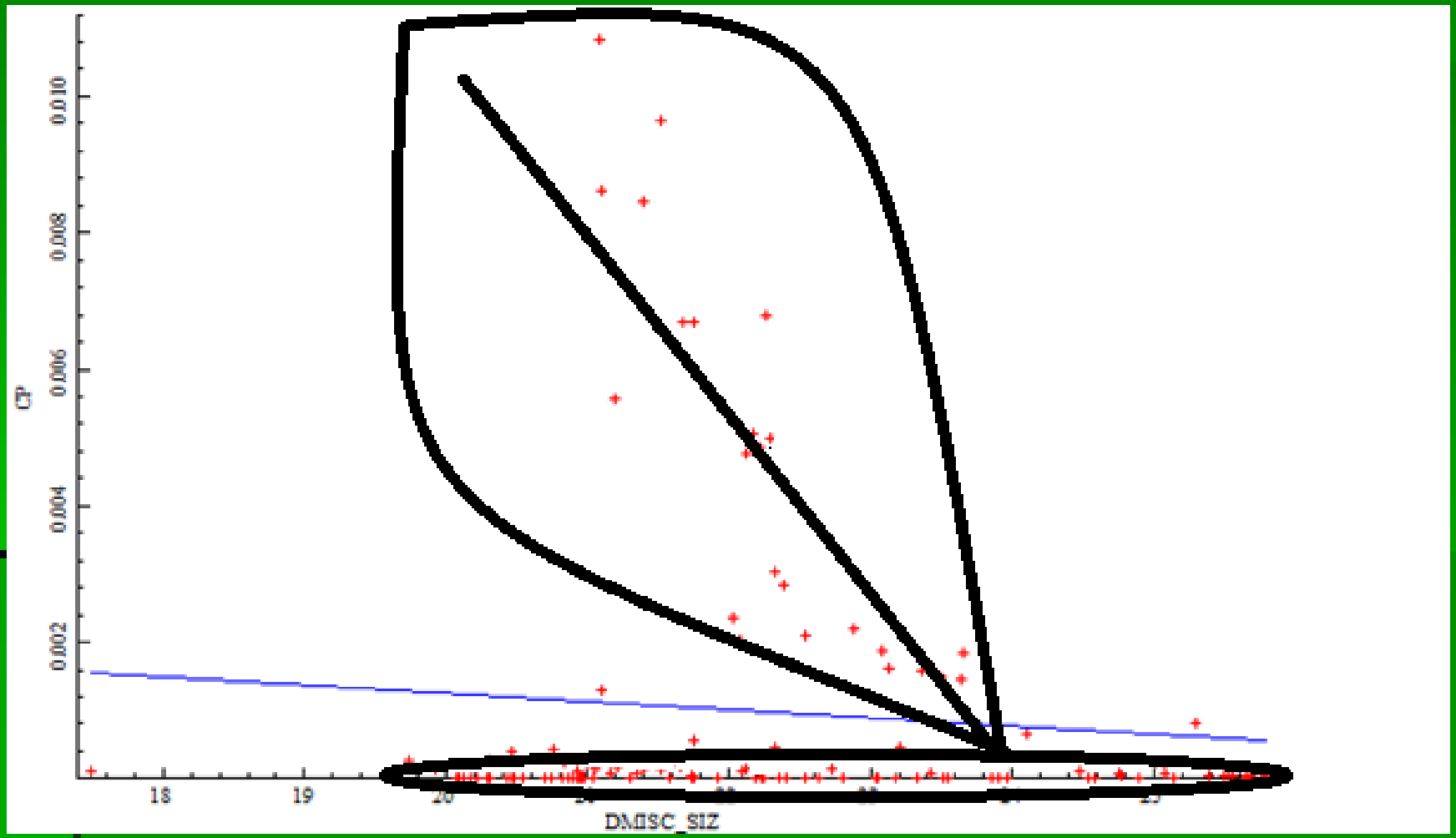
Firm size and CP: Chemical Sector



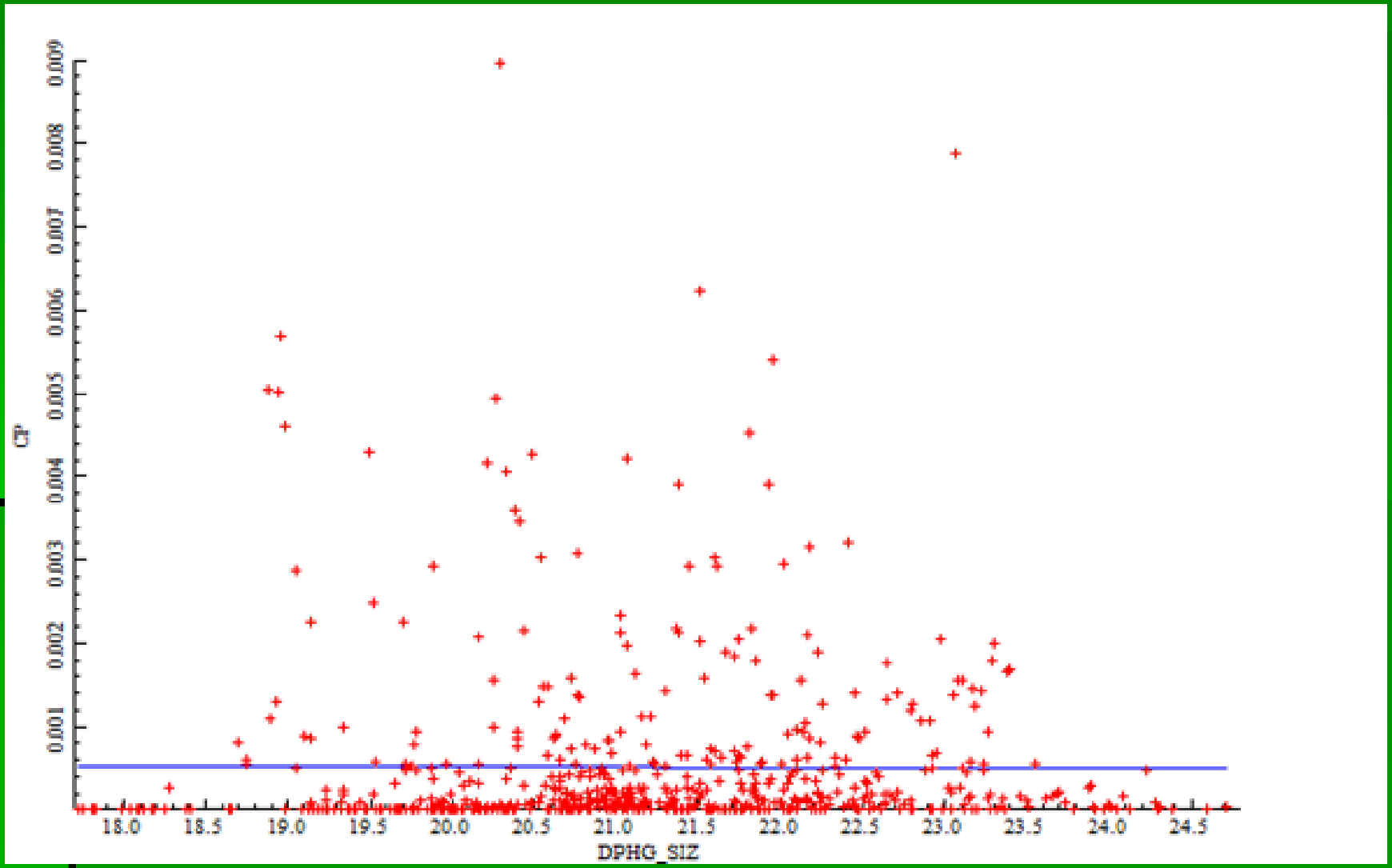
Firm size and CP: Construction Sector



Firm size and CP: Miscellaneous Sector

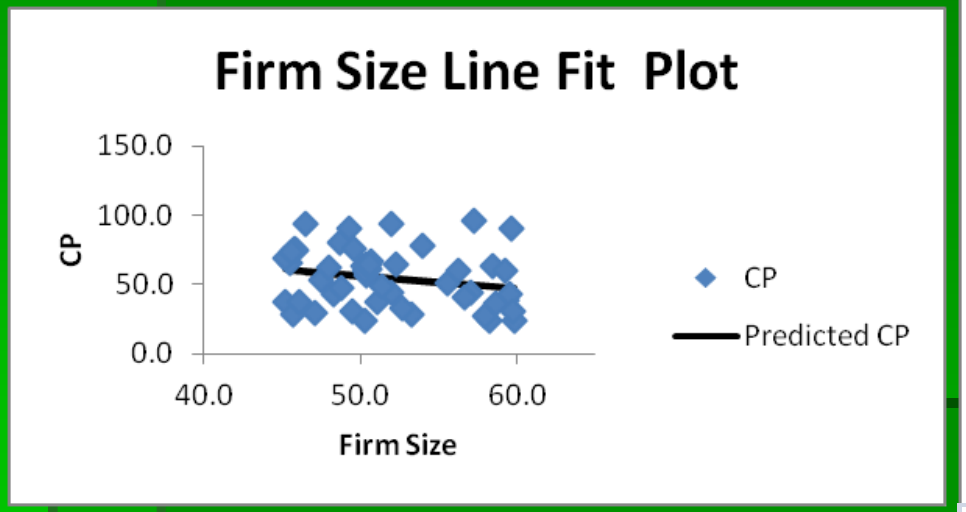


Firm size and CP: Personal Household Goods



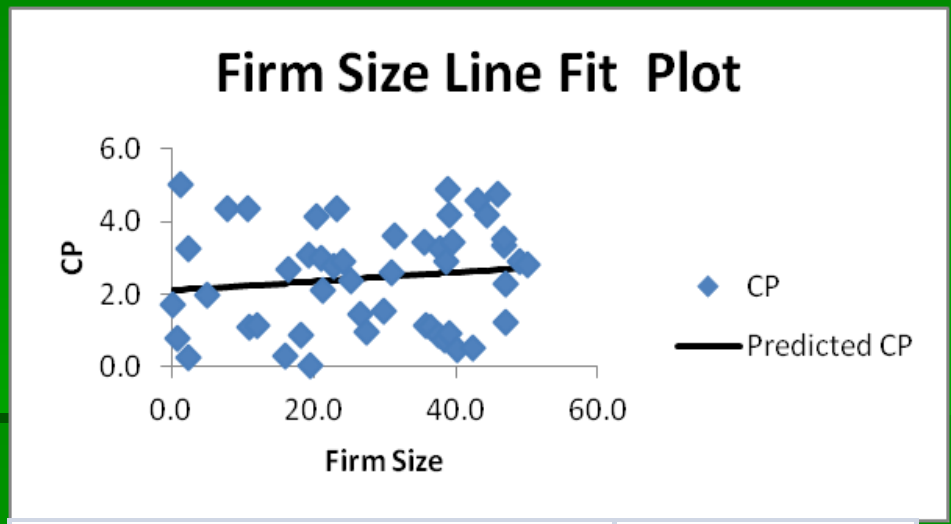
Sim: Financial & Services

R^23 = 0.046; SE = 20.5



	Coefficients	Standard Err	t Stat
Intercept	102.3625	32.12057	3.186819
Firm Size	-0.92502	0.616063	-1.5015

R^2 = 0.016; SE = 1.44

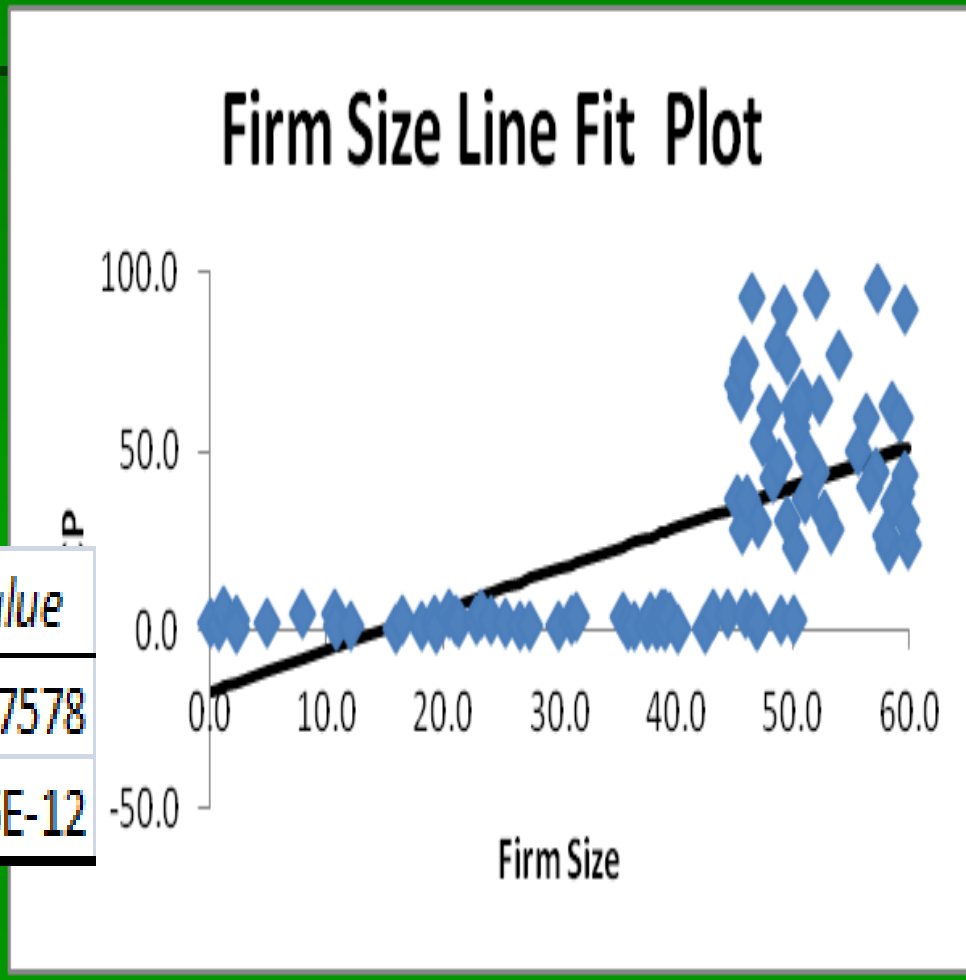


	Coefficients	Standard Err
Intercept	2.107138	0.439834
Firm Size	0.012355	0.013931

Combined Regression

Regression Statistics	
Multiple F	0.624346
R Square	0.389808
Adjusted I	0.383452
Standard E	23.48716
Observati	98

	Coefficients	*tandard Err*	*t Stat*	*P-value*
Intercept	-17.1269	6.278221	-2.72798	0.007578
Firm Size	1.140744	0.145667	7.831186	6.46E-12



Indy 500 Winning Speed vs YR

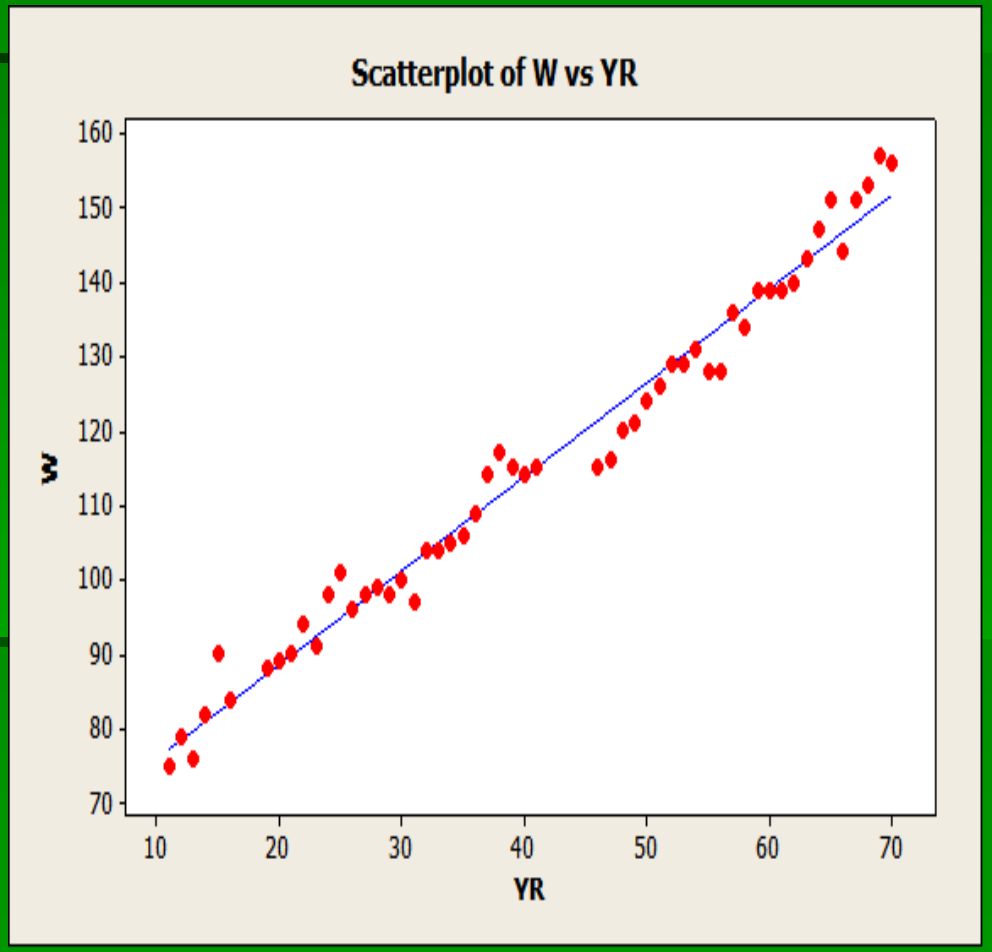
```
The regression equation is
W = 63.4 + 1.26 YR


Predictor      Coef  SE Coef      T      P
Constant     63.436    1.146  55.33  0.000
YR          1.25999  0.02561  49.20  0.000


S = 3.32674   R-Sq = 97.9%   R-Sq(adj) = 97.9%
```

b = 1.26 WinSpeed increases by 1.26 mph/year.

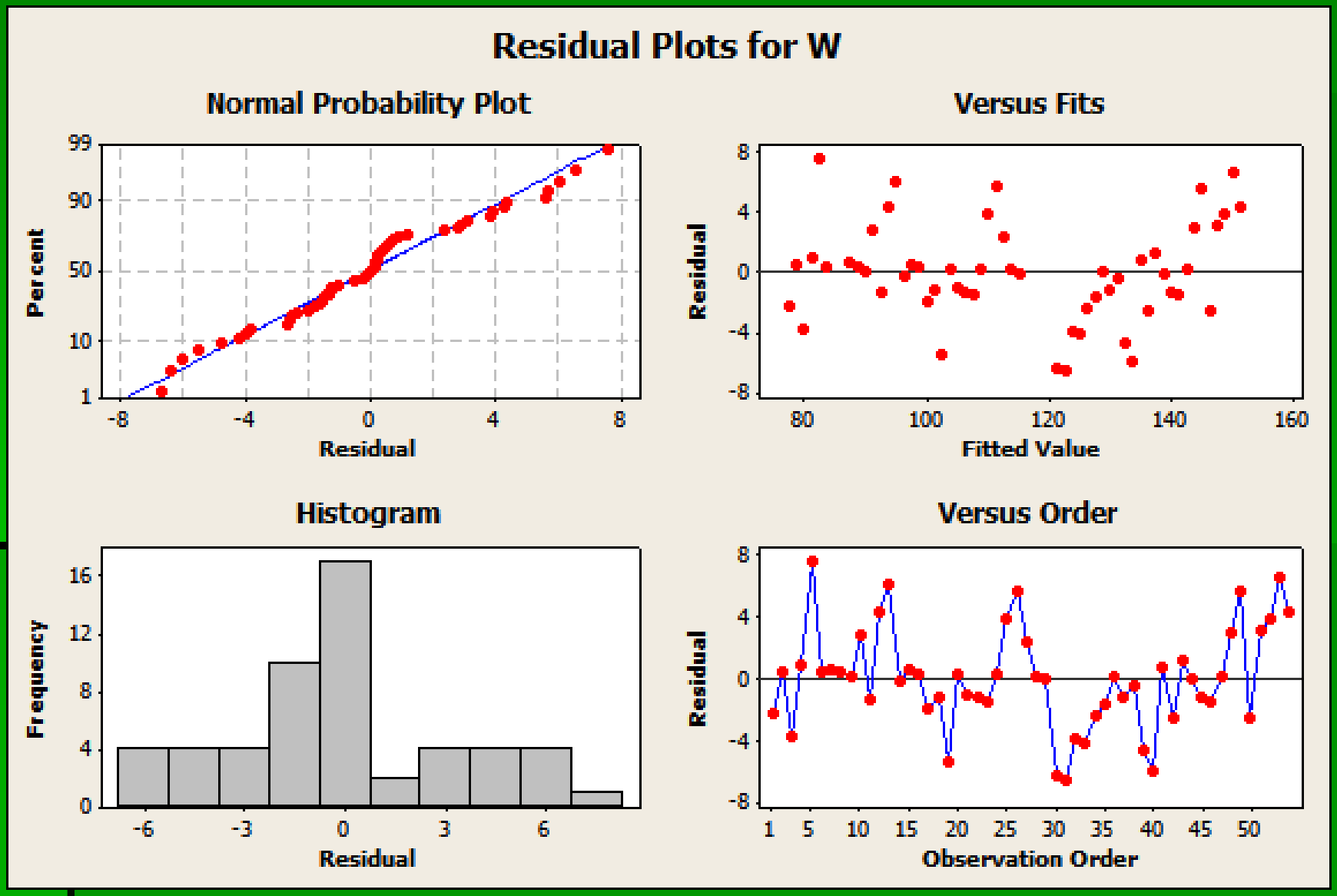
a = 63.4 the winning speed in 1900 –

SE = 3.33 accuracy of fit of the regression line.

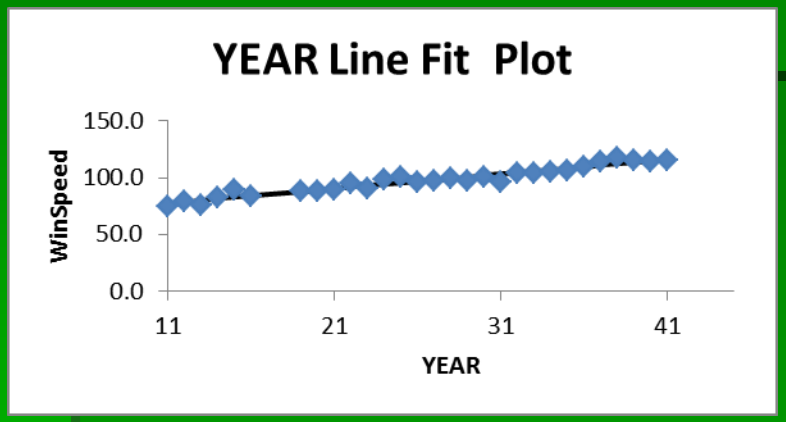
2/3 of the data (63.4 + 1.26 T) ± SE

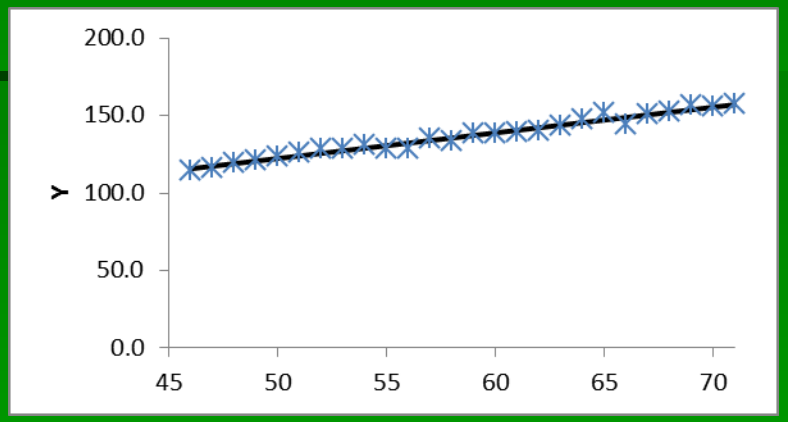


Residuals from Indy 500



Pre & Post WW2 Fits





R Square	0.939596
Adjusted I	0.937359
Standard I	2.998946

	Coefficients	*andard Err*	*t Stat*
Intercept	63.62633	1.744782	36.46663
YEAR	1.2746	0.062195	20.49368

	Coefficients	*andard Err*	*t Stat*
Intercept	39.71655	3.33018	11.92625
X Variable	1.658242	0.056464	29.36814

R Square	0.972927
Adjusted I	0.971799
Standard I	2.159333

Residual Plots – Pre & Post