CHR Solutions

Regression 101 Simplified

Kent Larsen
CHR Solutions

CHRSolutions.com

Regression Analysis - Defined

- Statistical process for estimating the relationships among variables.
- Focus is on the relationship between a dependent variable and one or more independent variables.
- Helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed.
- Less commonly, the focus is on a quantile, or other location parameter of the conditional distribution of the dependent variable given the independent variables.

Regression Analysis - Defined

- Widely used for prediction and forecasting
- Also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships.
- In restricted circumstances, regression analysis can be used to infer causal relationships between the independent and dependent variables.
  - Can lead to illusions or false relationships,
  - Correlation does not imply causation.

Classical Assumptions - Regression Analysis :

- The sample is representative of the population for the inference prediction.
- The error is a random variable with a mean of zero conditional on the explanatory variables.
- The independent variables are measured with no error. (Note: If this is not so, modeling may be done instead using errors-in-variables model techniques).
- The predictors are linearly independent, i.e. it is not possible to express any predictor as a linear combination of the others.
- The errors are uncorrelated, that is, the variance–covariance matrix of the errors is diagonal and each non-zero element is the variance of the error.
- The variance of the error is constant across observations (homoscedasticity). If not, weighted least squares or other methods might instead be used.

Classic Examples

- Personal Income – Auto Sales
- Age = Insurance Claims
- Interest rates = Economic Activity

No One EVER Said:

Personal Income – Auto Sales

- If your income is not what I predict, you can't by a car

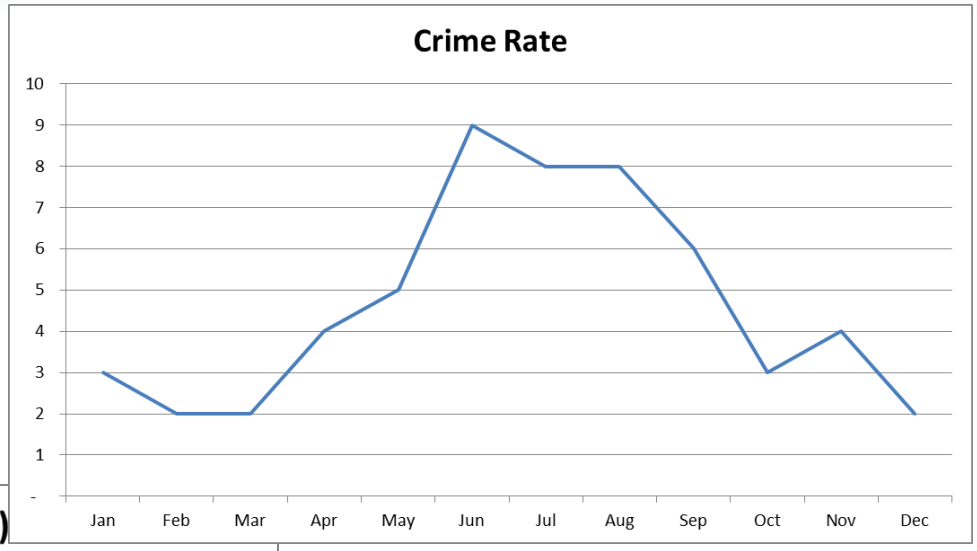
Age = Insurance Claims

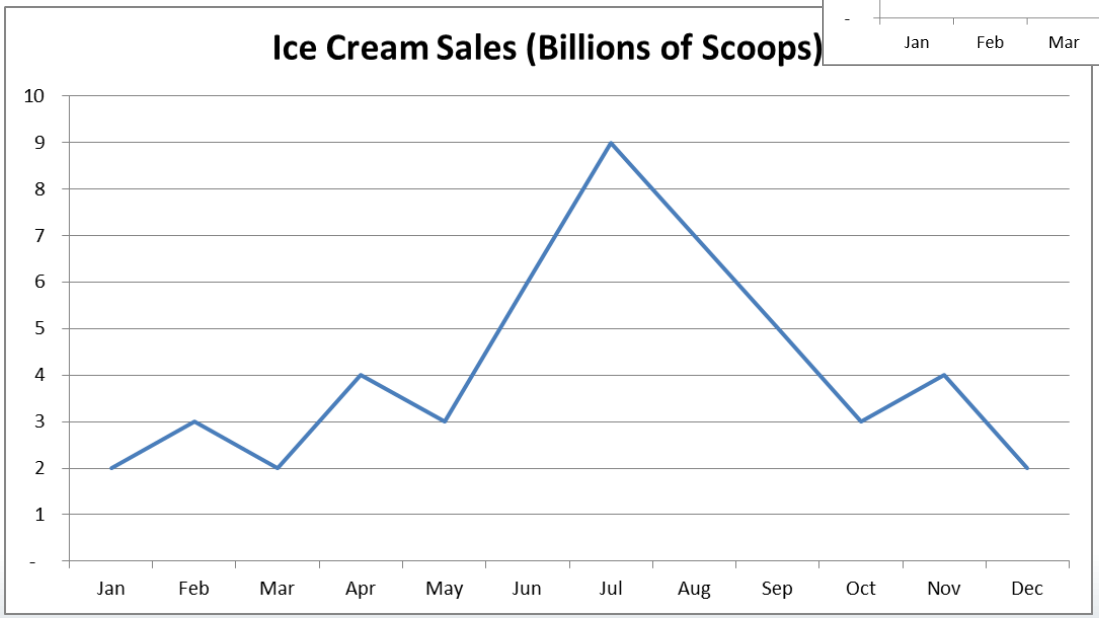
- If you are young, you can't make as many claims as if you were old

Interest rates = Economic Activity

- If interest rates are high, you can't invest

Correlation = Causation?





Example

- Do access line counts predict total cost

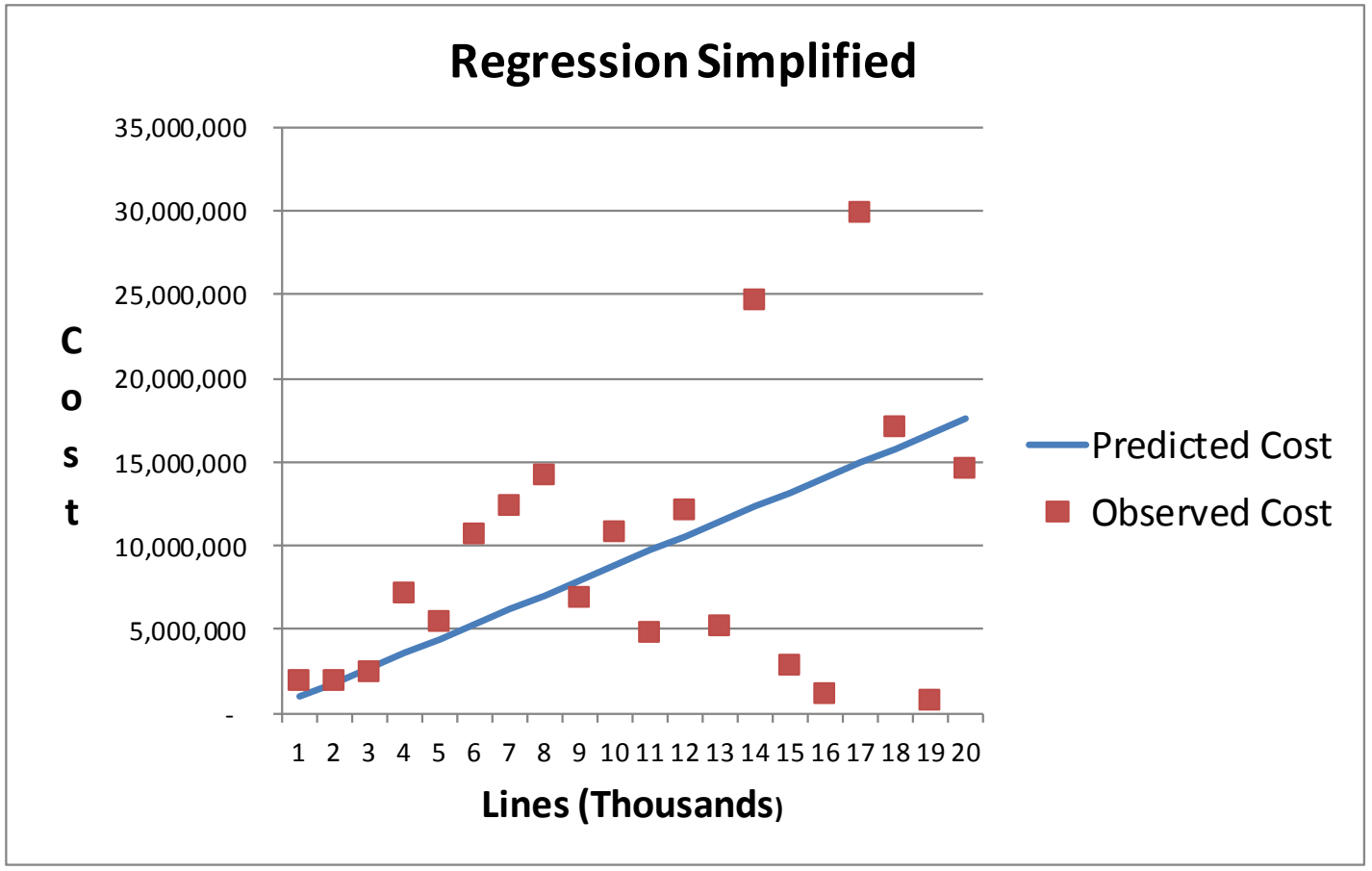
FYI – the “poor man’s regresssion tool” is the Trend Line function within Excel graphs.

Left click on chart “series”, right click, “Add Trenline”

Examples

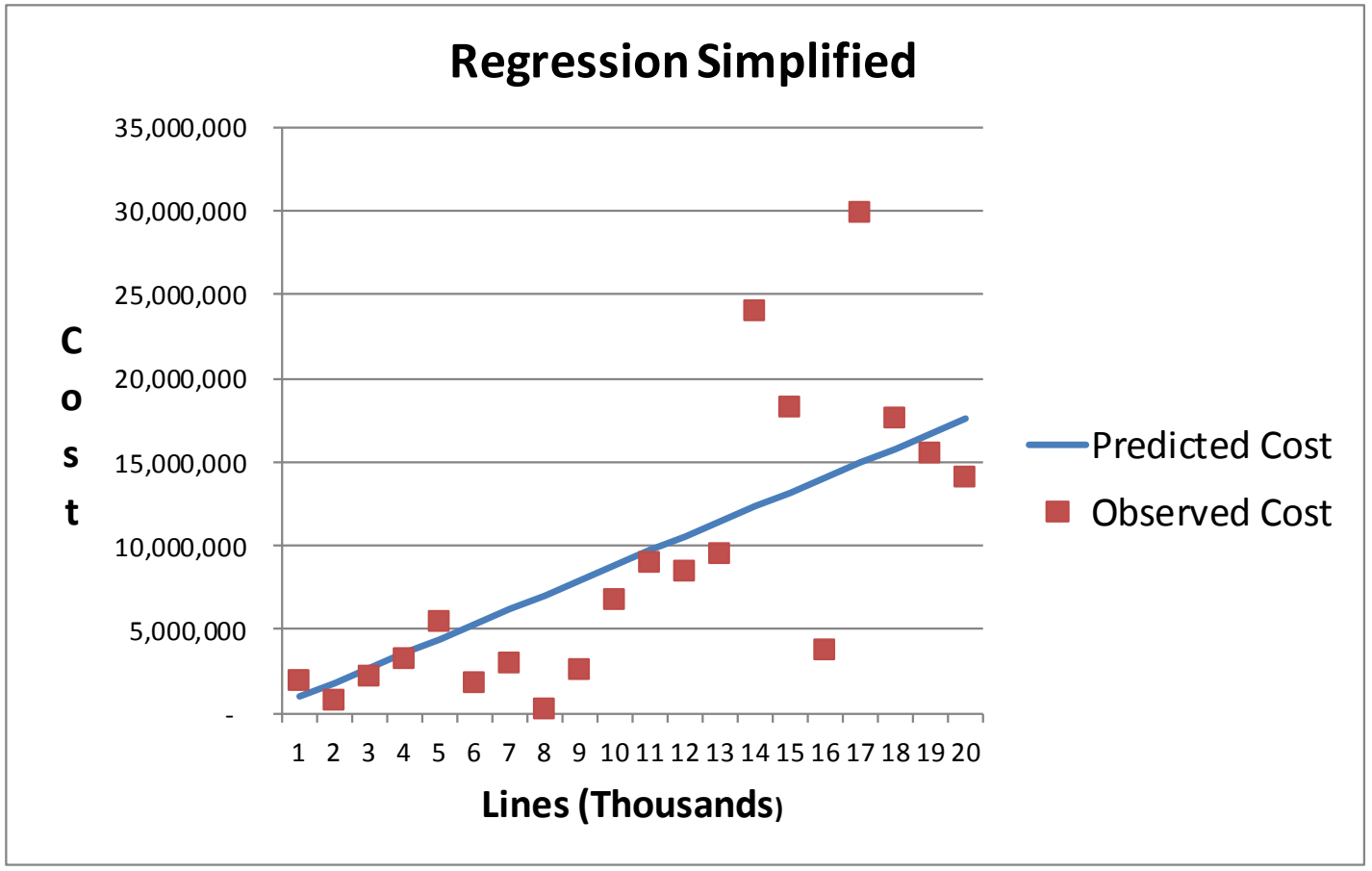
- **Input**
  - Lines
  - Actual (Observed) Total Costs
- **Output**
  - Constant = a = $80,000
  - Line Variable = b = $875
- **Cost = a + (b x Lines)**

Wide Variance



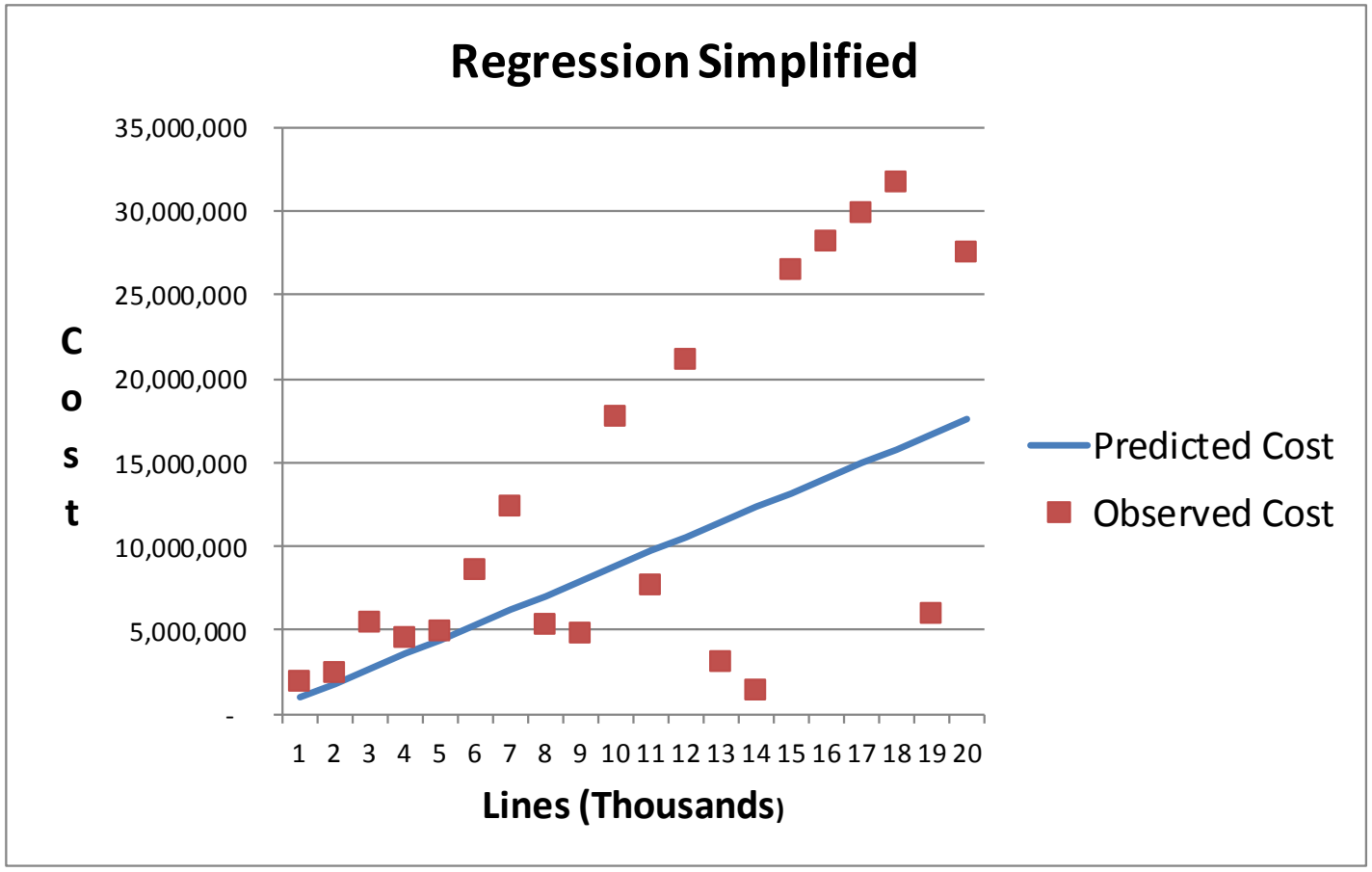
Average	9,335,906
Median	7,011,343

Variance on High



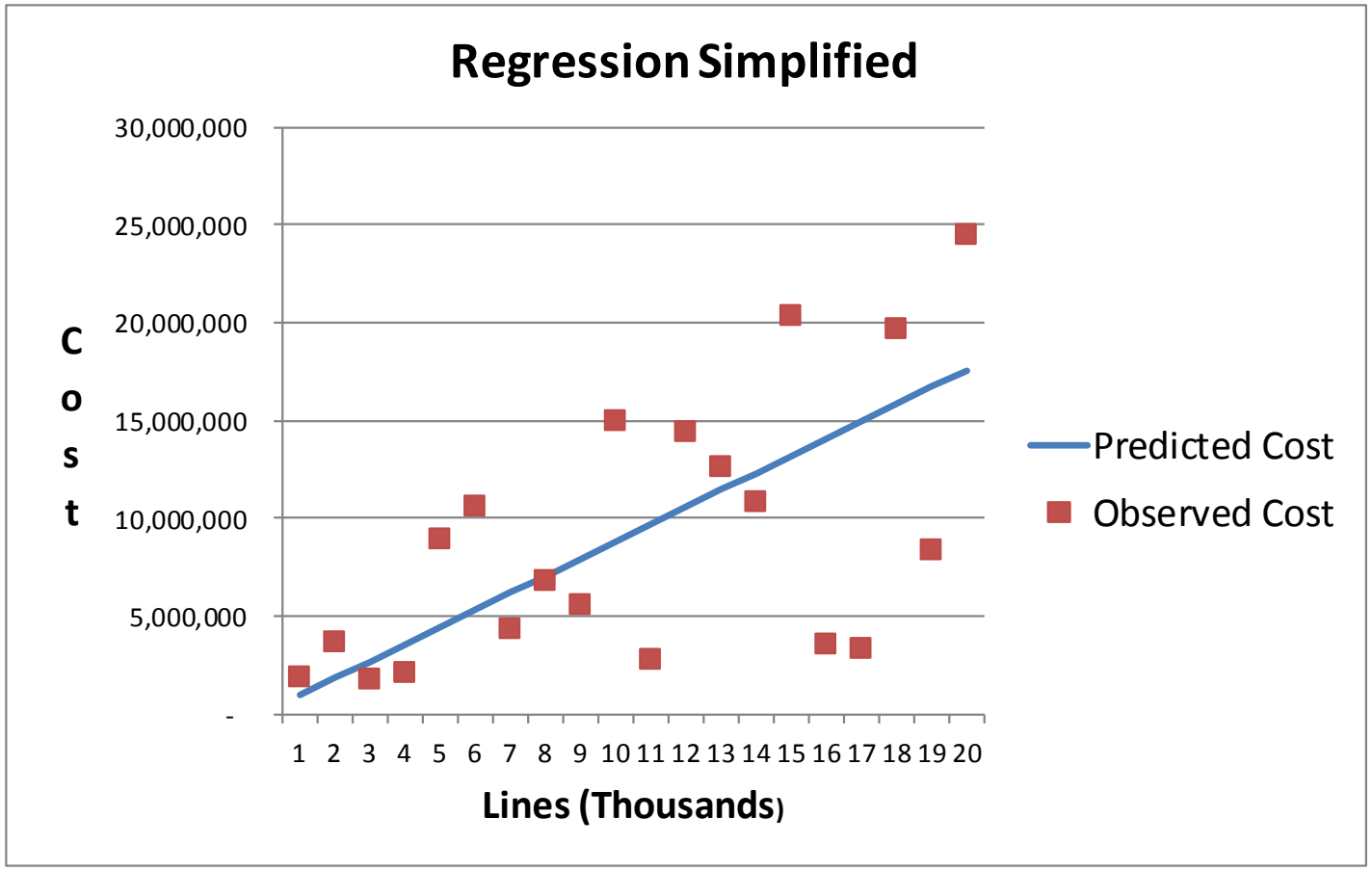
Average	8,909,444
Median	6,122,985

Fit on Low – Low Confidence for Linear Relationship



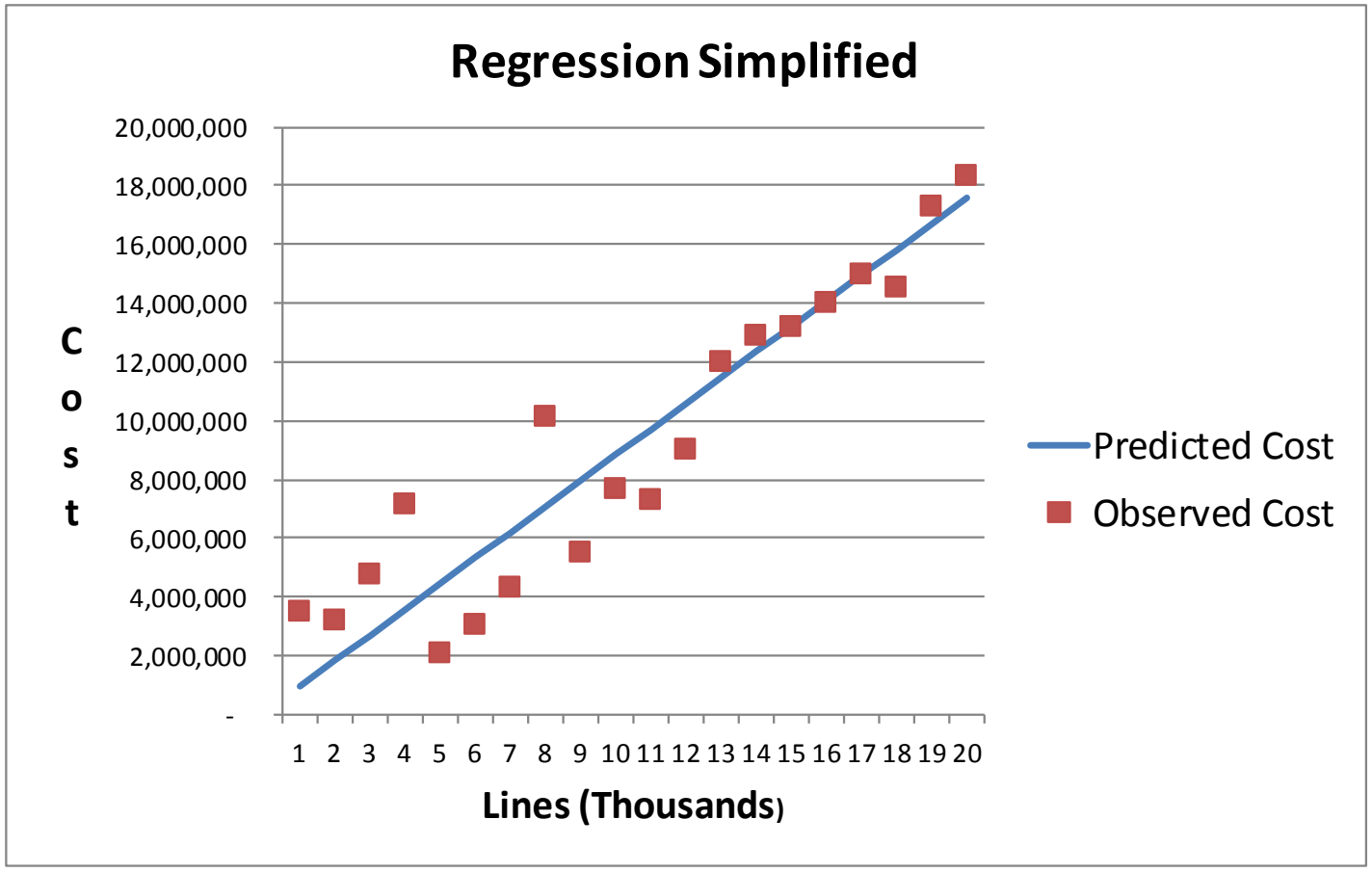
Average	12,547,685
Median	6,855,003

Consistent and Wide Variance – Low Confidence



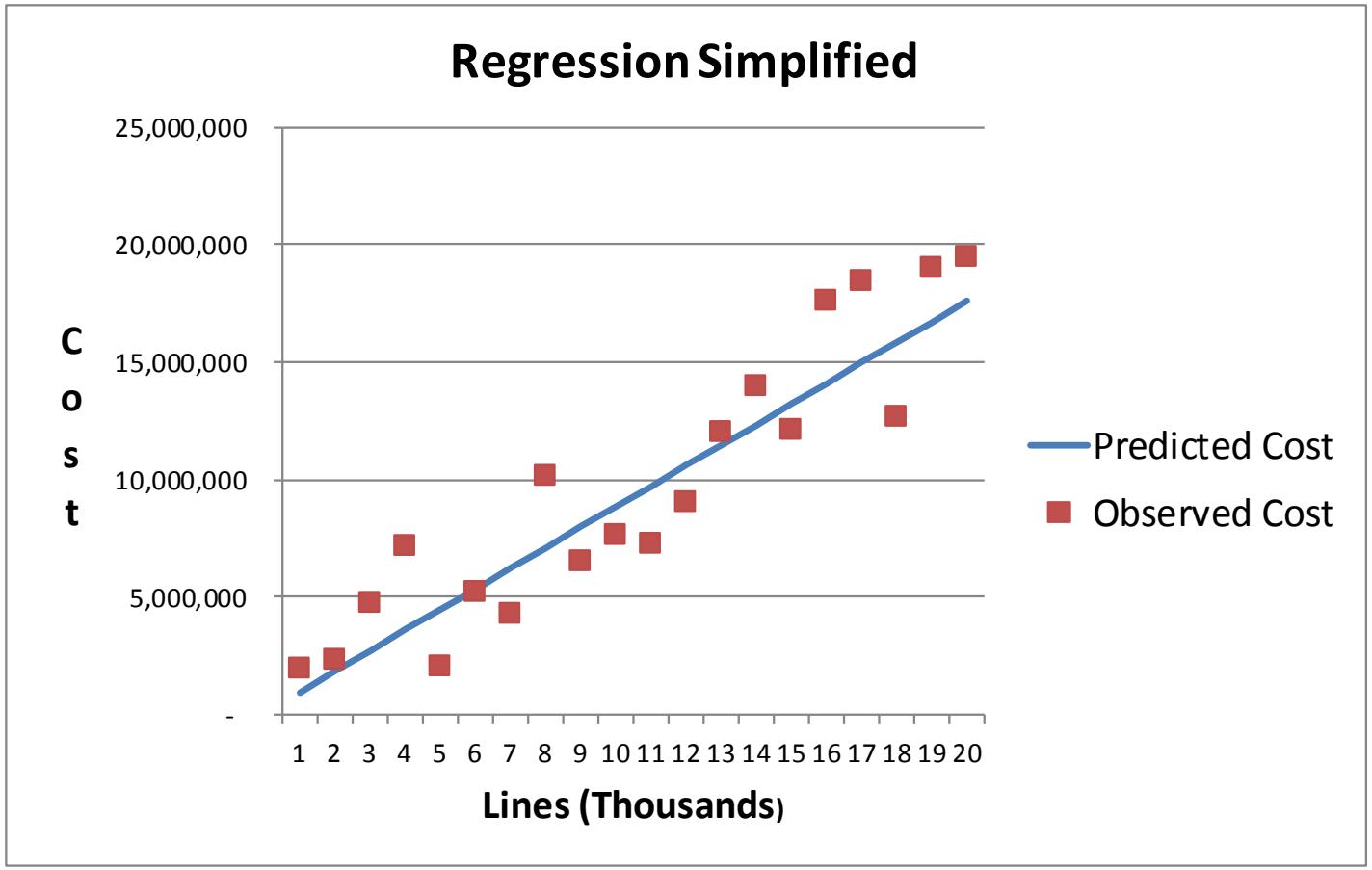
Average	9,045,770
Median	7,592,342

Fit on High – Economies of Scale



Average	9,243,699
Median	8,330,000

Low Variance Across Spectrum Increases Confidence



Average	9,681,703
Median	8,330,000

FCC Used

- Loops
- Road Miles
- Road Crossings
- % Undepreciated Plant
- Density
- Exchanges
- Bedrock
- Difficulty
- Climate
- Tribal Land
- Park Land
- Urban
- Alaska, Midwest, Northeast

Quantile Regression…EXPLAINED!!

- Quantile regression is desired if conditional quantile functions are of interest. One advantage of quantile regression, relative to the ordinary least squares regression, is that the quantile regression estimates are more robust against outliers in the response measurements. However, the main attraction of quantile regression goes beyond that. In practice we often prefer using different measures of central tendency and statistical dispersion to obtain a more comprehensive analysis of the relationship between variables.
- In ecology, quantile regression has been proposed and used as a way to discover more useful predictive relationships between variables in cases where there is no relationship or only a weak relationship between the means of such variables. The need for and success of quantile regression in ecology has been attributed to the complexity of interactions between different factors leading to data with unequal variation of one variable for different ranges of another variable.
- Another application of quantile regression is in the areas of growth charts, where percentile curves are commonly used to screen for abnormal growth.
- The mathematical forms arising from quantile regression are distinct from those arising in the method of least squares. The method of least squares leads to a consideration of problems in an inner product space, involving projection onto subspaces, and thus the problem of minimizing the squared errors can be reduced to a problem in numerical linear algebra. Quantile regression does not have this structure, and instead leads to problems in linear programming that can be solved by the simplex method. The fact that the algorithms of linear programming appear more esoteric to some users may explain partially why quantile regression has not been as widely used as the method of least squares.

Global Warming and USF Costs…EXPLAINED!!

- Quantile regression is desired if conditional quantile functions are of interest. One advantage of quantile regression, relative to the ordinary least squares regression, is that the quantile regression estimates are more robust against outliers in the response measurements. However, the main attraction of quantile regression goes beyond that. In practice we often prefer using different measures of central tendency and statistical dispersion to obtain a more comprehensive analysis of the relationship between variables.

- In **ecology**, quantile regression has been **proposed and used** as a way to discover more useful predictive relationships between variables **in cases where there is no relationship or only a weak relationship** between the means of such variables. The need for and success of quantile regression in ecology has been attributed to the complexity of interactions between different factors leading to data with unequal variation of one variable for different ranges of another variable.

- Another application of quantile regression is in the areas of growth charts, where percentile curves are commonly used to screen for abnormal growth.
- The mathematical forms arising from quantile regression are distinct from those arising in the method of least squares. The method of least squares leads to a consideration of problems in an inner product space, involving projection onto subspaces, and thus the problem of minimizing the squared errors can be reduced to a problem in numerical linear algebra. Quantile regression does not have this structure, and instead leads to problems in linear programming that can be solved by the simplex method. The fact that the algorithms of linear programming appear more esoteric to some users may explain partially why quantile regression has not been as widely used as the method of least squares.

So What

- FCC implemented quantile regression as the new standard for judging telco costs.
- Primary and supplemental explanations for methods, assumptions and exceptions
- Ultimately decided that the Top 10% of costs were ineligible for recovery
  - Limited to predictive model rather than actual costs
- Ultimately capped @ $250 – NOT Regression

First Round

- Fail: Multiple measures, e.g., several discrete capital and operating expense regressions
- Fail: limit cost recovery and excluded from recycled support
- Fail: no recognition of substitute technology or operating strategy
- Fail: the “Similarly Situated” fraud

Fail – Substitute Technologies

- Companies chose microwave rather than fiber to the node
- Model predicted company could have spent several times what it did spend for fiber
- Model predicted company should not have spent what it did spend for microwave
- Company SAVED USF millions (avoided FTTn) in favor of cheaper microwave

Fail – Operational Discretion

- Companies chose to maintain current network rather than invest in FTTx
- Operating expenses higher than model predicted
- Capital expenditures lower
- Company booted for operating expense “Fail”
- Company SAVED USF millions

Fail – No Actual Discovery of “Waste”

- Companies that could be judged “wasteful” yet under caps untouched
- Companies that are extremely small, startups (last 20 years), served huge territories judged “wasteful”.
- Holding Companies Conspicuously Under-represented

Fail – Similarly Situated

- **No distinction for ownership, governance**
  - Coops, Commercials, Holding Companies
- **Kansas**
  - Of 34 LECS
  - 16 reduced CPL by average of $448
- **Southeast**
  - Of 115 LECS
  - 22 reduced CPL by average of $166
- **Small**
  - Of 65 < 500 lines
  - 13 reduced CPL by average of $985

Should vs. True

Should

- **Used to indicate obligation, duty, or correctness, typically when criticizing someone's actions:**
  - "You should have been careful and invested less in your network".
- **Indicating a desirable or expected state:**
  - "By now USF funding should be less".

True

- **In accordance with fact or reality:**
  - "The accounting records reflect true costs".

Should vs. True

- **FCC no longer believes:**
  - Certified Public Accountants
  - Telco Accountants
  - Consultants
  - Part 32 / 36 / 64 / 65 / 69
  - You or Me or Its own rules
- **FCC now believes:**
  - Statisticians
  - Gi-Normous public databases
  - Proprietary Analytical Software
- **Fundamentally, a Lazy Approach**

Irony = Reality

- The total funding $$ cap on USF HCL remains

Andy Rooney Moment:

“Didja ever notice that AT&T and Verizon, et al. are not in this fight?”

Shape-Shifter

- Every time a new regression analysis is performed…

- New Top 10
- New assumption of predicted costs

- If LECs Tend to Reduce Costs
  - You better run faster than the average
- If LECs Tend to Increase Costs
  - You better run at about, or slightly slower, than the average
- Regardless…there is only so much cash to go around

Fini