Stat 107: Introduction to Business and Financial Statistics

Class 14: Review of Simple Regression

Fun! Science! Facts!

EVERYDAY MYSTERIES

Fun Science Facts from the Library of Congress

<< HOME   << See More Everyday Mysteries>>   << Ask a Question >>

Find [ ] in Everyday Mysteries Pages GO

Question:

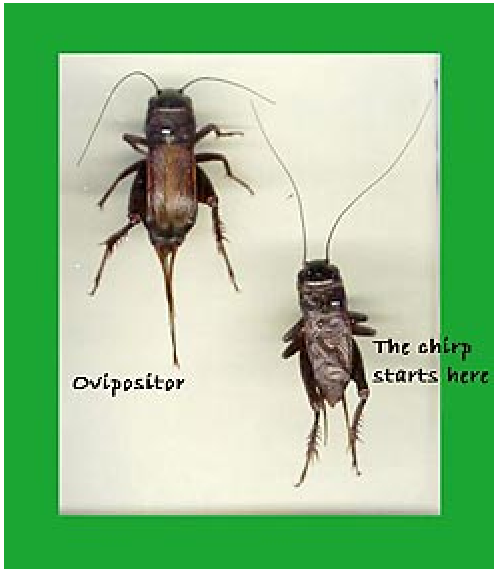
Can you tell the temperature by listening to the chirping of a cricket?

Answer:

Yes!

The frequency of chirping varies according to temperature. To get a rough estimate of the temperature in degrees fahrenheit, count the number of chirps in 15 seconds and then add 37. The number you get will be an approximation of the outside temperature.

So, how do crickets make that chirping sound?



Female (left) an Male (right) crickets

Where did this rule come from?

The frequency of chirping varies according to temperature. To get a rough estimate of the temperature in degrees fahrenheit, count the number of chirps in 15 seconds and then add 37. The number you get will be an approximation of the outside temperature.

- They fit a line to the data set. This is what regression does-it relates a Y variable to an X variable.
- ***There are many ways to fit a line to data***, though one method is the most popular (but not always the best method).

Mathematically they are saying Temp = 37+Chirps…how wrong are they????

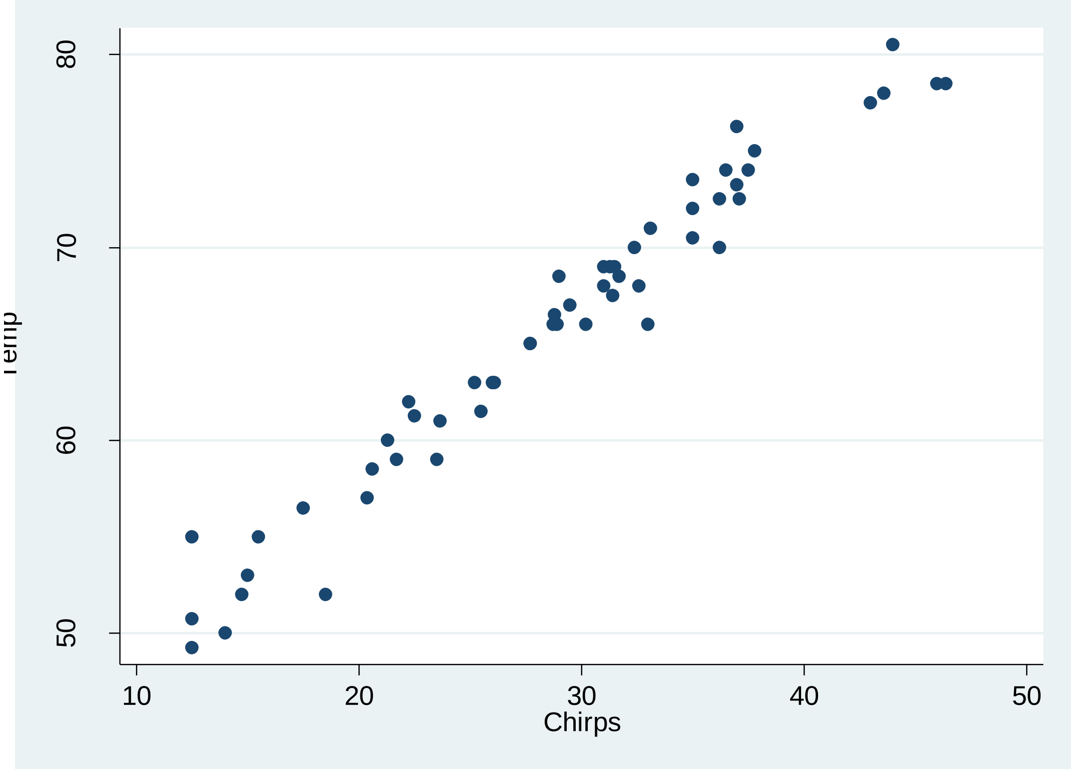
A Cricket Video

- Slightly creepy youtube video

https://www.youtube.com/watch?v=VAkX_yEa1Os fname="http://www.people.fas.harvard.edu/~mparzen/stat107/chirp_data.csv"

Cricket Data

- X= number of chirps per 15 seconds
- Y = Temperature



Data from http://blog.globe.gov/sciblog/2007/10/05/measuring-temperature-using-crickets/

Fitting Line Method 1

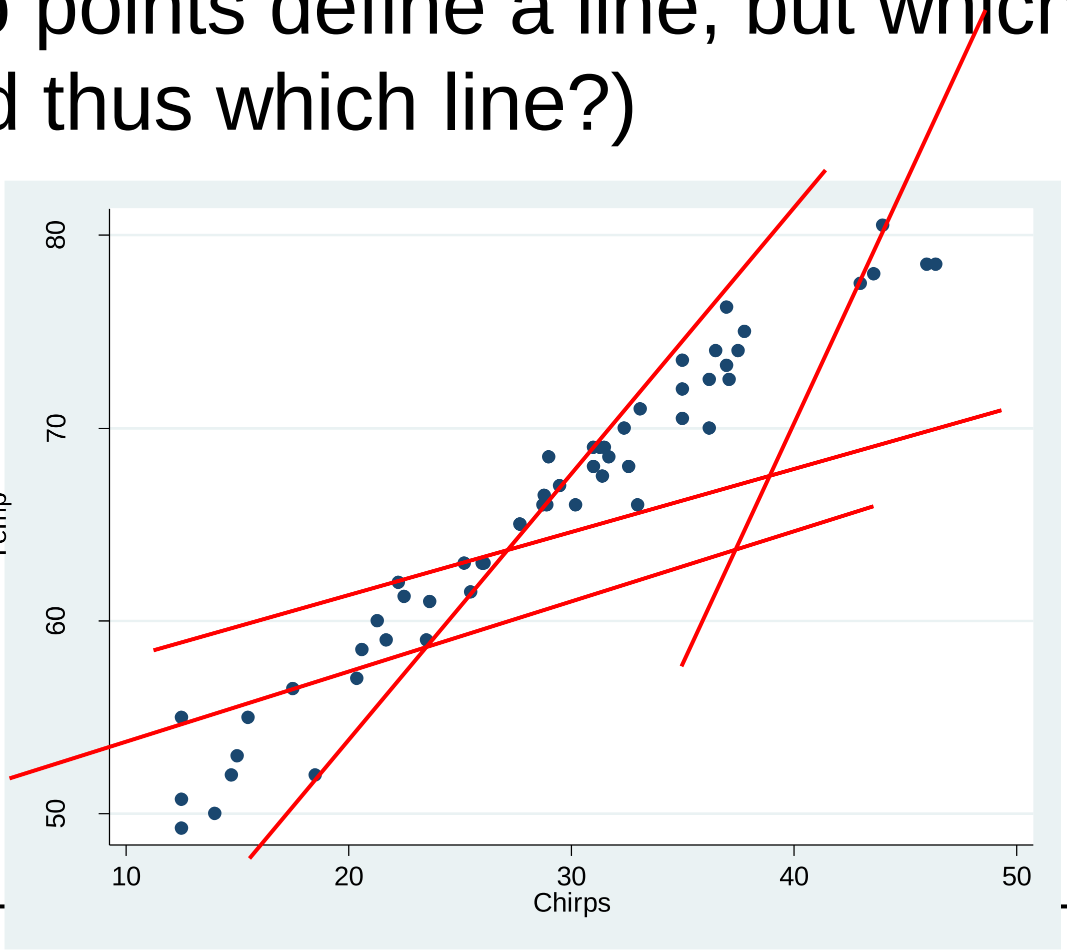
- Draw a line by hand
- Not exactly scientific..not **algorithmic**

> al·go·rithm
> /ˈalgəˌriT͟Həm/
>
> Noun
> A process or set of rules to be followed in calculations or other problem-solving operations, esp. by a computer.

- **We want easily reproducible results.**

Which Two Points?

- Two points define a line, but which two points (and thus which line?)



Fitting Line Method 2

- Two points define a line….so we need two points.
- Split the X axis in two, so there is a lower half and upper half group of points.
- Find two “good” points in each group
- Fit a line connecting these two points.
- I just made this method up, by the way.

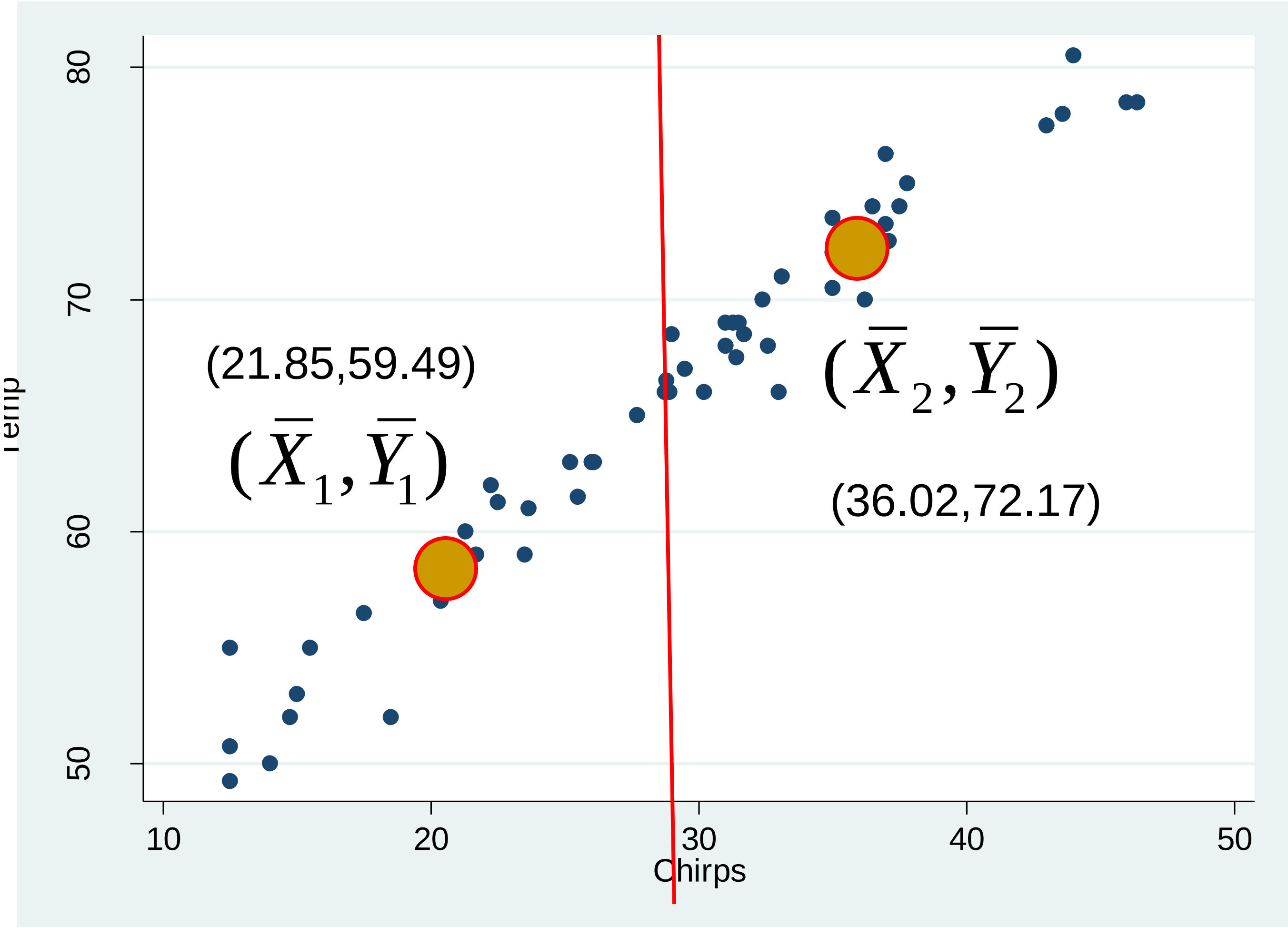
Make This an Algorithm

- Calculate the median of the X's
- Separate the data into two groups
- Find $(\bar{X}_1, \bar{Y}_1)$ and $(\bar{X}_2, \bar{Y}_2)$
- Calculate the line between these two points
- Recall the equation of a line formula

$$Y - Y_1 = m(X - X_1)$$

$$m = \frac{Y_2 - Y_1}{X_2 - X_1}$$

A Picture



29.5

The Fitted Line

$$m = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{(72.17 - 59.49)}{(36.02 - 21.85)} = 0.90$$

$$Y - Y_1 = m(X - X_1) \Rightarrow Y = 39.8 + .9X$$

So for prediction, we say Temp = 39.8+.9(Chirps)

Interpret the Line

- How do we interpret this:

$$Temp = 39.8 + 0.9(Chirps)$$

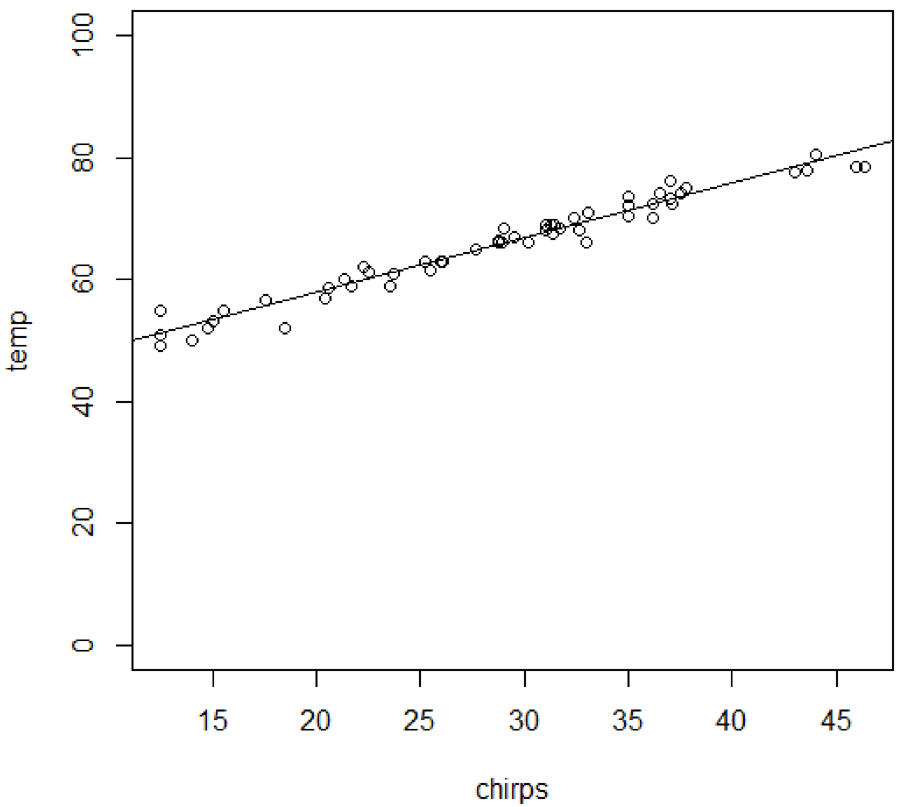
- If chirps goes up by 1 unit (1 additional chirp per time period), temp goes up by 0.9.
- How wrong are we???

R Code

```
mikeline=function(x,y) {

    xmed=median(x)
    xbar1=mean(x[x<=xmed])
    ybar1=mean(y[x<=xmed])
    xbar2=mean(x[x>xmed])
    ybar2=mean(y[x>xmed])

    slope = (ybar2-ybar1)/(xbar2-xbar1)
    inter = ybar1-slope*xbar1
    cat("Intercept = ",inter," Slope = ",slope,"\n")
    ##return(slope)
}

    > mikeline(chirps,temp)
    Intercept =  39.9446  Slope =  0.8945971
    >
```

The Fitted Line Plot

```
plot(chirps,temp,ylim=c(0,100))
abline(39.944,.894)
```



Pause: The Equation of a Line

English words for the French word montant

amount, figure, rising, sum

- Most Americans have been brainwashed

$$Y = mX + b$$

- (allegedly in France they use y=sx+b)

- As adults, we will now use the notation

$$Y = b_0 + b_1 X$$

https://www.math.duke.edu//education/webfeats/Slope/Slopederiv.html

Notation for **Our** Line

- We need to be able to distinguish between our observed Y values, and the Y values that our line produces.

- So given a slope and intercept, we produce what is called the fitted line:

$$\hat{Y}_i = b_0 + b_1 X_i$$

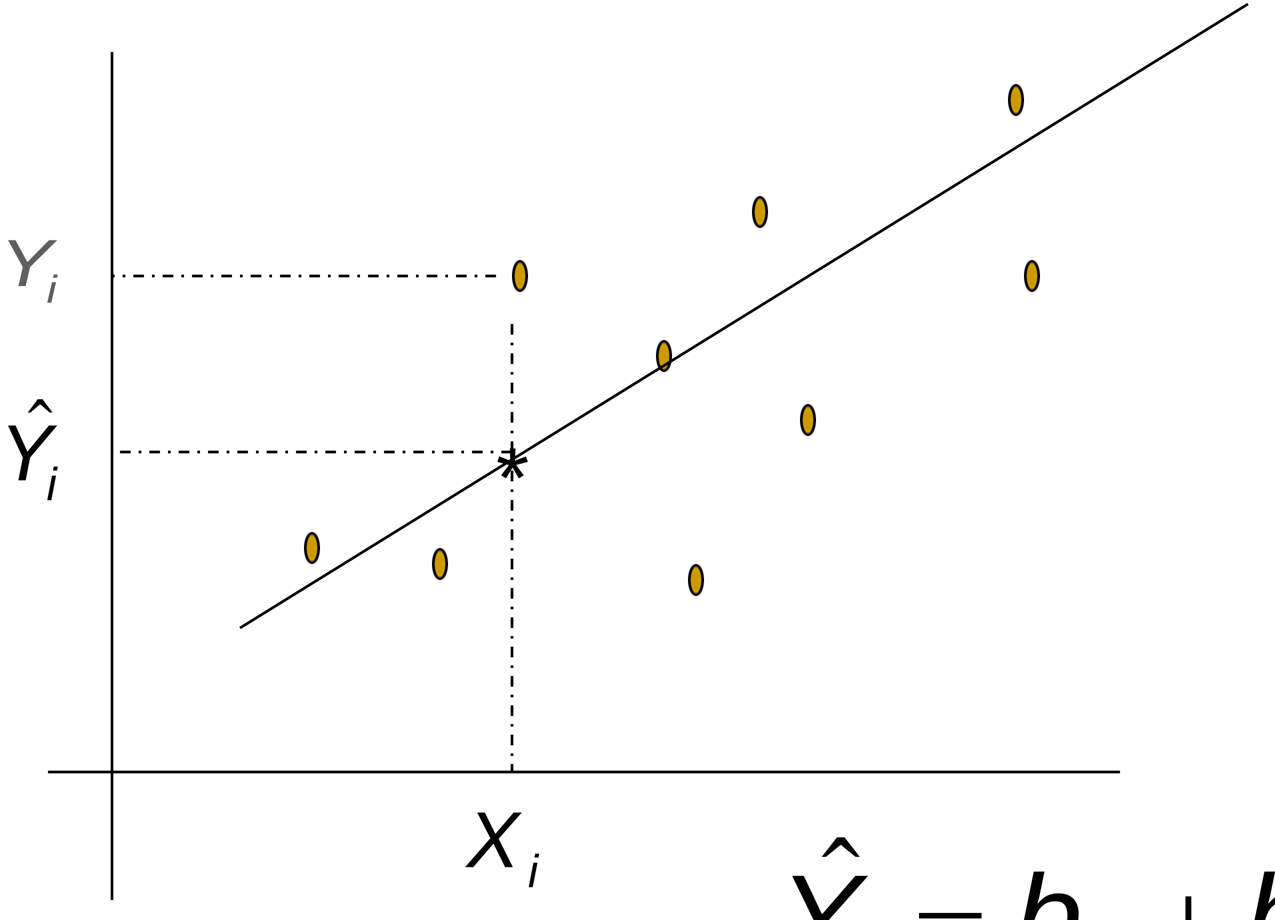
Pause: To Fit a Line to Data

- Fitting a line to data means to find "good" values of b_0 and b_1 .
- We define our fitting error as

$$e_i = Y_i - b_0 - b_1 X_i = Y_i - \hat{Y}_i$$

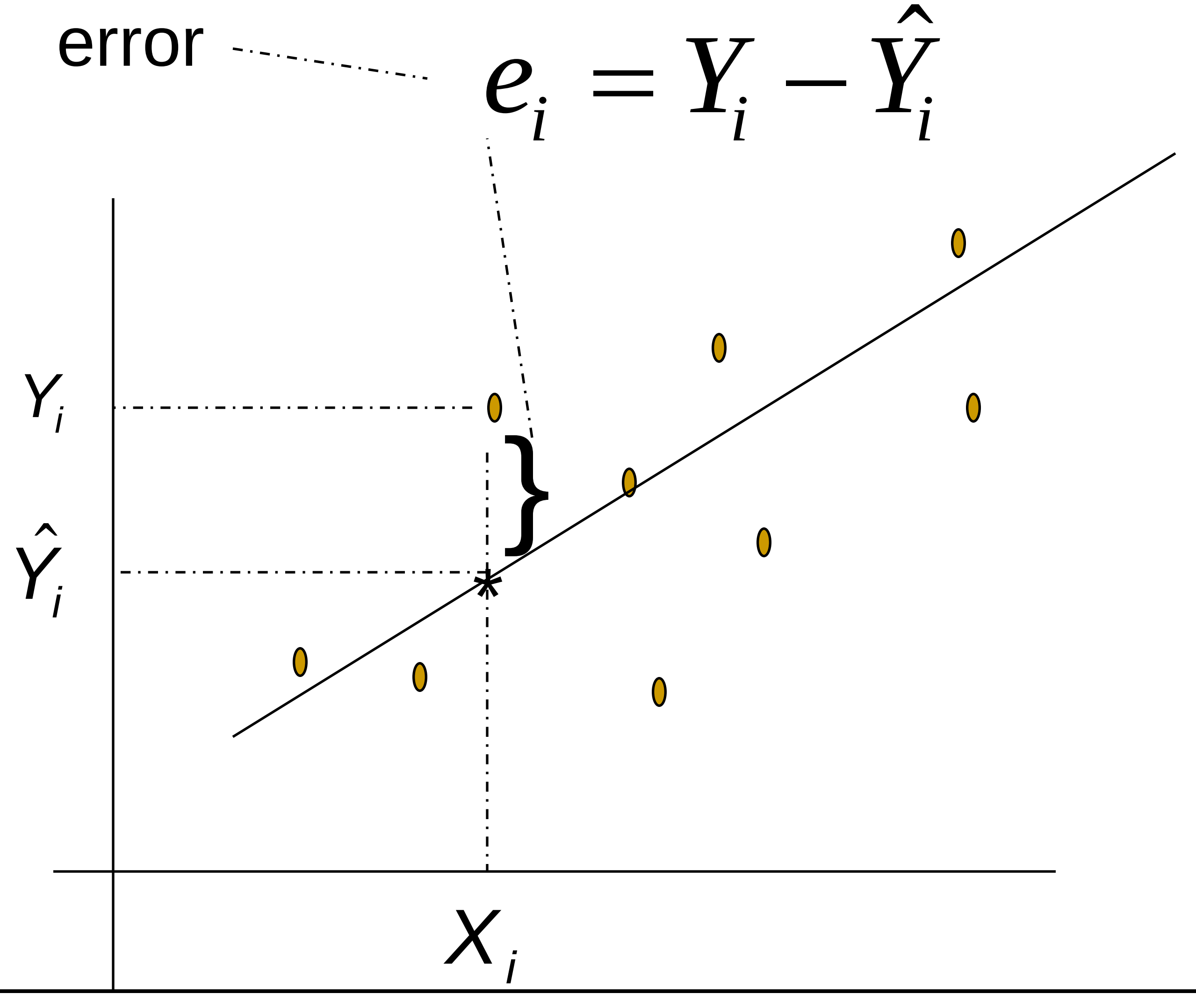
- Ideally, we want all the errors to be zero. Is this always possible?
- So we need a **criterion function**

Observed versus Fitted Values



$$\hat{Y}_i = b_0 + b_1 X_i$$

The Errors



19

Criterion Function 1(line method 3)

- What if we decide to find b_0 and b_1 that satisfy the criterion

$$\underset{b_0, b_1}{find} \quad \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) = \sum_{i=1}^{n} e_i = 0$$

- That is, if we can't make each error zero, lets make the sum of them zero.
- This method can work ok, but also has some problems.
- We solve it using optimization methods.

Solver in Excel

- Excel is a necessary evil for any undergrad to learn.
- Almost every nonprofit, government, consulting, I-banking, etc.. job requires knowledge of it.
- Solver is an Excel routine that can find function minimums, as well as solve f(x)=0.
- More details are available on the course website.

Using Excel

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)$$

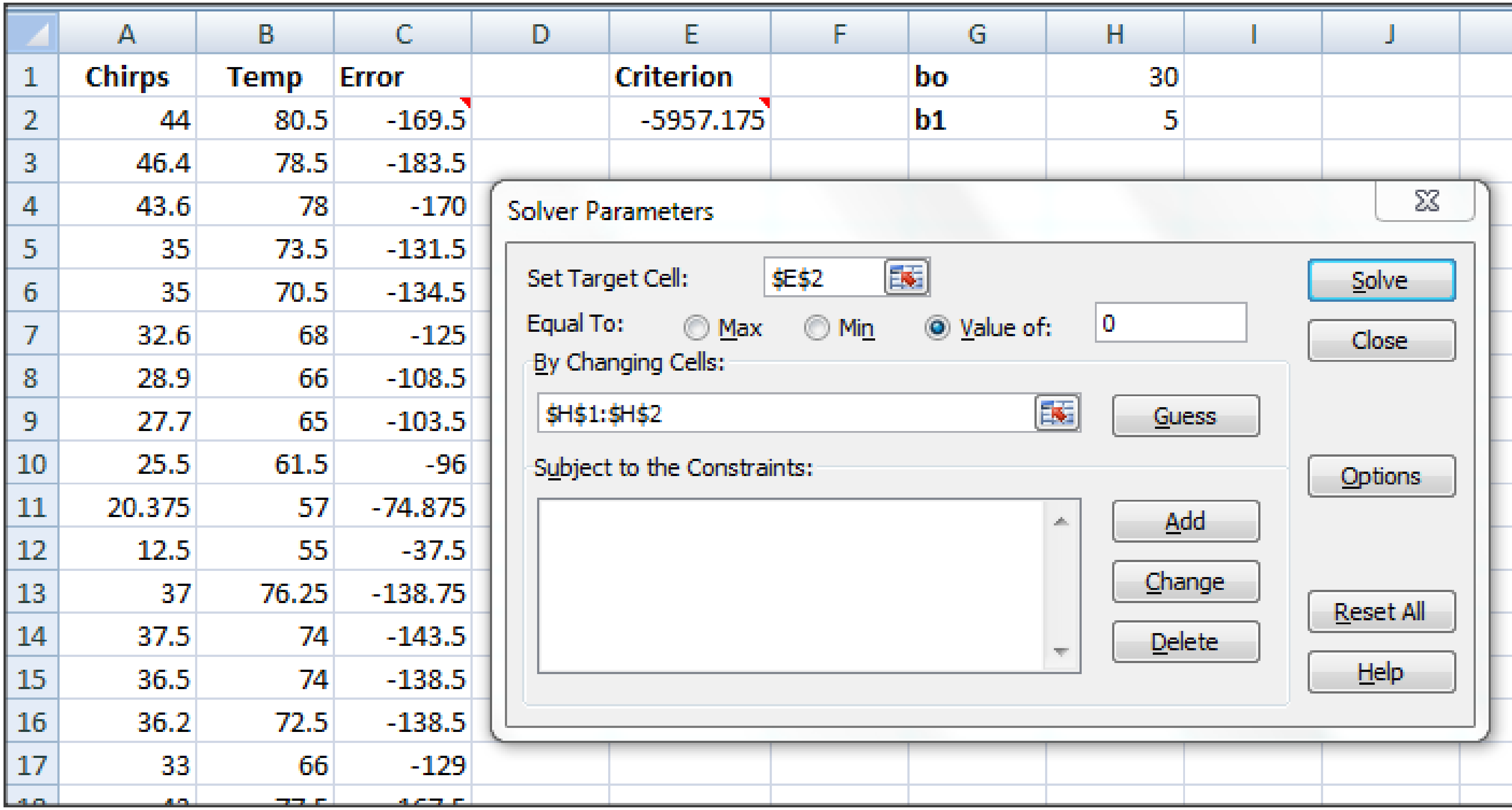
	A	B	C	D	E	F	G	H
1	Chirps	Temp	Error		Criterion		bo	30
2	44	80.5	-169.5		-5957.175		b1	5
3	46.4	78.5	-183.5					
4	43.6	78	-170					
5	35	73.5	-131.5					
6	35	70.5	-134.5					
7	32.6	68	-125					
8	28.9	66	-108.5					
9	27.7	65	-103.5					

SUM(C2:C56)

(B2-H$1-H$2*A2)

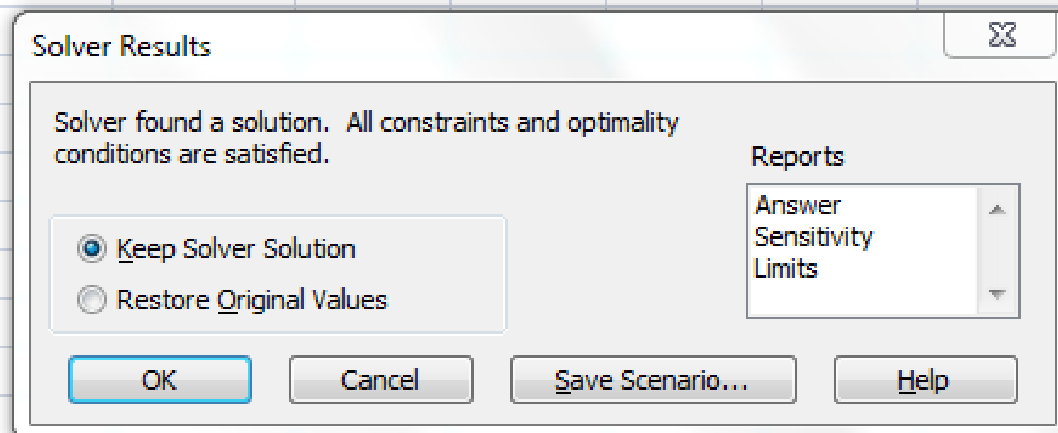
$$(Y_i - b_0 - b_1 X_i)$$

Using Solver in Excel



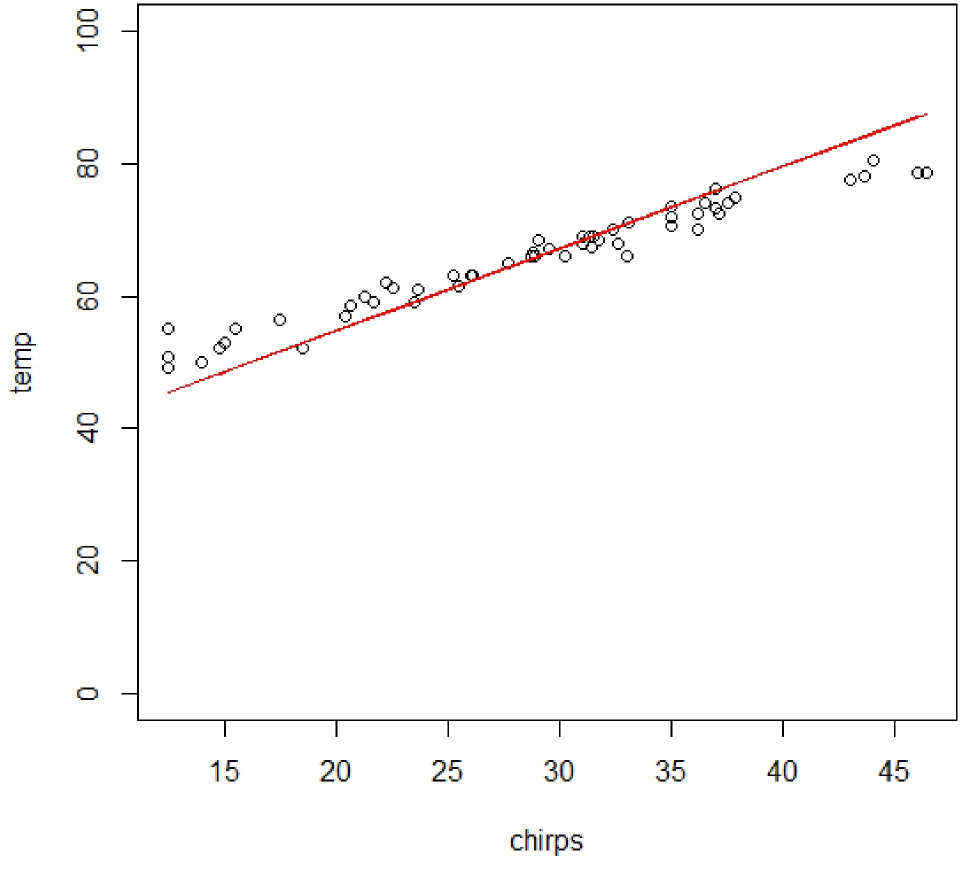
The Solution

	A	B	C	D	E	F	G	H
1	Chirps	Temp	Error		Criterion		bo	29.86962
2	44	80.5	-4.12146		1E-06		b1	1.24436
3	46.4	78.5	-9.10792					
4	43.6	78	-6.12372					
5	35	73.5	0.077779					
6	35	70.5	-2.92222					
7	32.6	68	-2.43576					
8	28.9	66	0.168376					
9	27.7	65	0.661608					
10	25.5	61.5	-0.1008					
11	20.375	57	1.776545					
12	12.5	55	9.57588					
13	37	76.25	0.339059					
14	37.5	74	-2.53312					
15	36.5	74	-1.28876					



The Resulting Graph

Temp = 29.87 + 1.24*Chirps

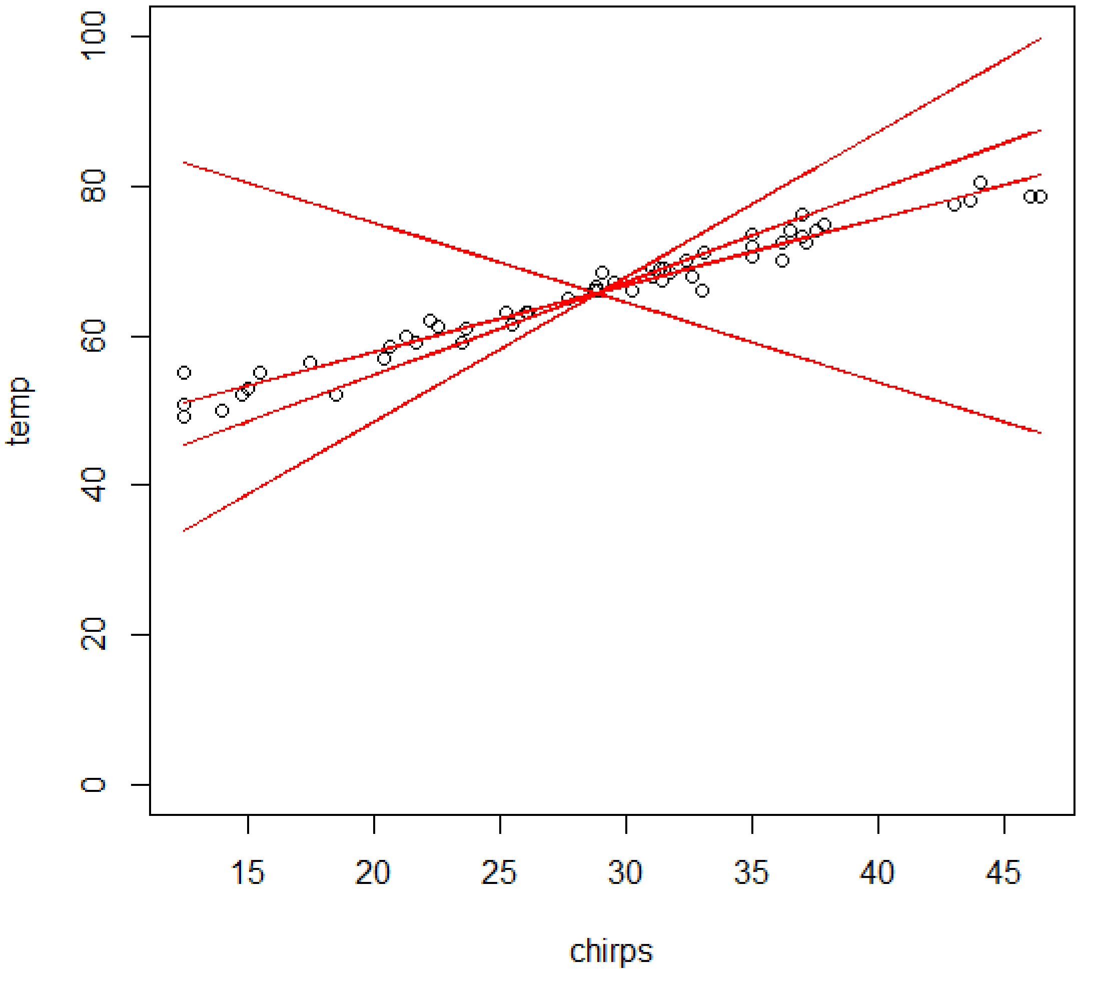


A Problem

- If we change our initial starting values, we get a different solution.
- The solution is not unique(!).
- On the next slide we show visually several different solutions, each obtained by using different starting values in Solver.

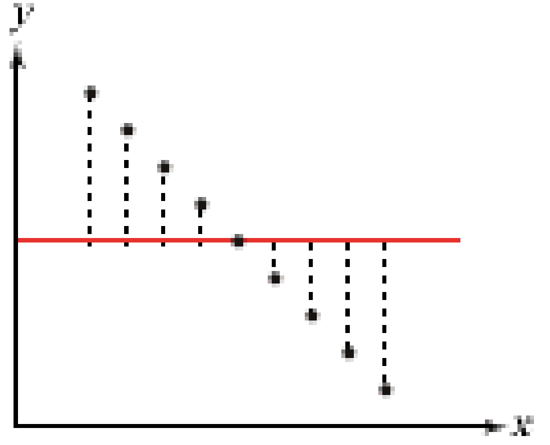
Different Valid Solutions

ALL these solutions make the sum of the errors zero!



Not a Good Criterion Function

- Consider the following graph with fitted line using the current criterion function.



- A good line of fit should follow the general direction of the points for a given set of data- **this criterion function doesn’t guarantee a good fit.**

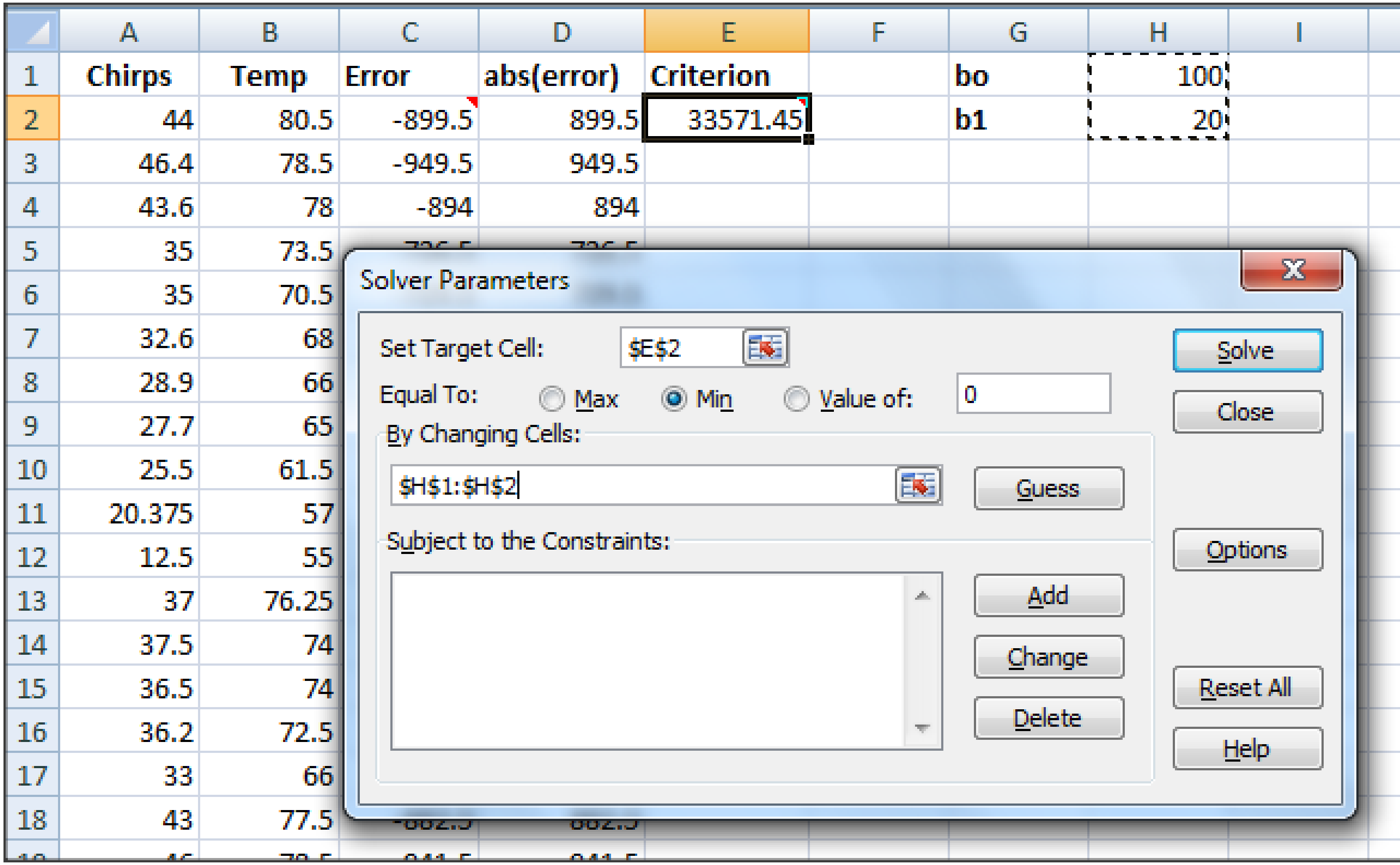
Criterion Function 2(line method 4)

- Consider the line found by solving the following Criterion function

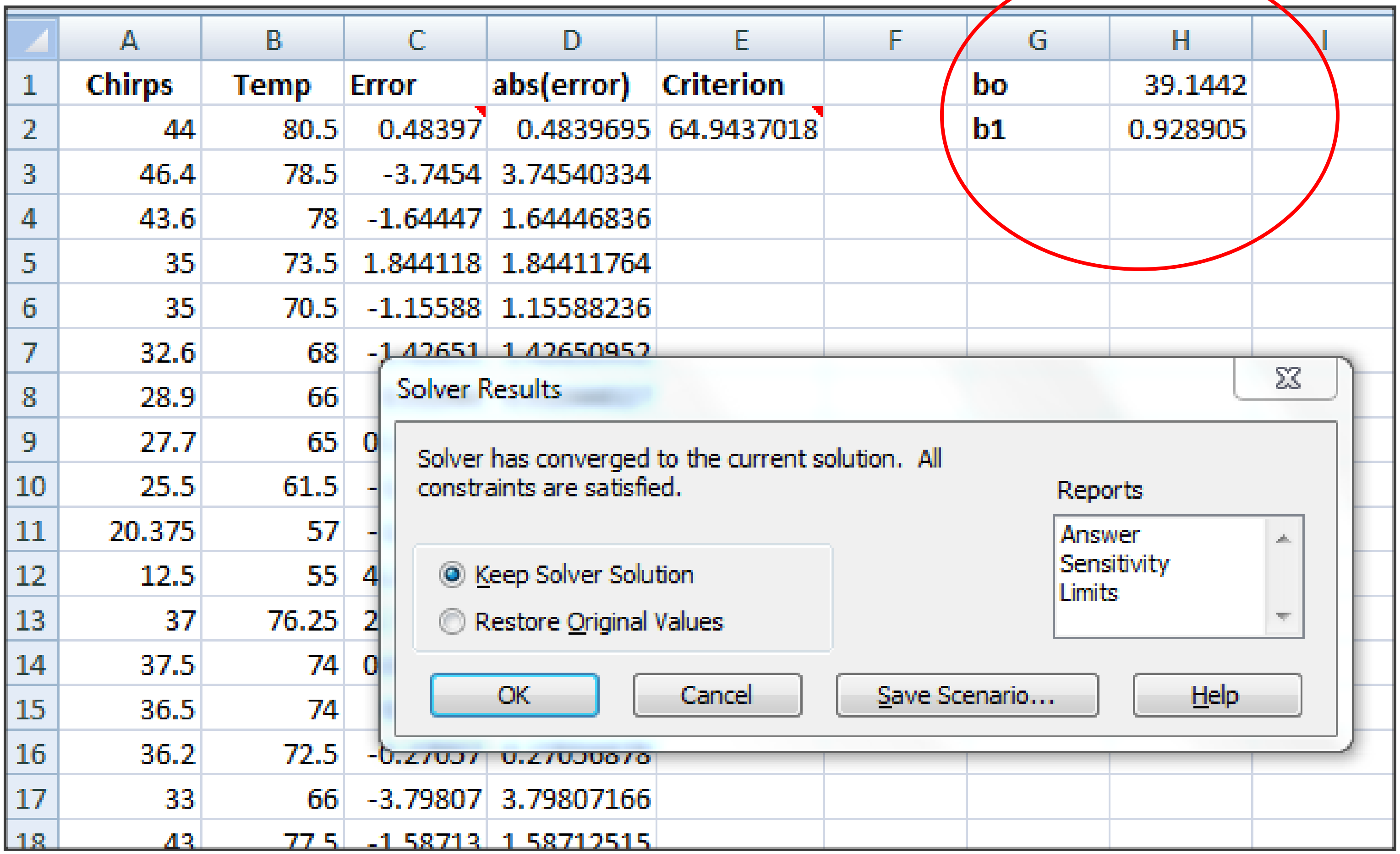
$$\min_{b_0, b_1} \sum_{i=1}^{n} | Y_i - b_0 - b_1 X_i |$$

- This is called Least Absolute Deviation
- Can we use Calculus to solve this?
- Back to Solver.

Setting Up Solver



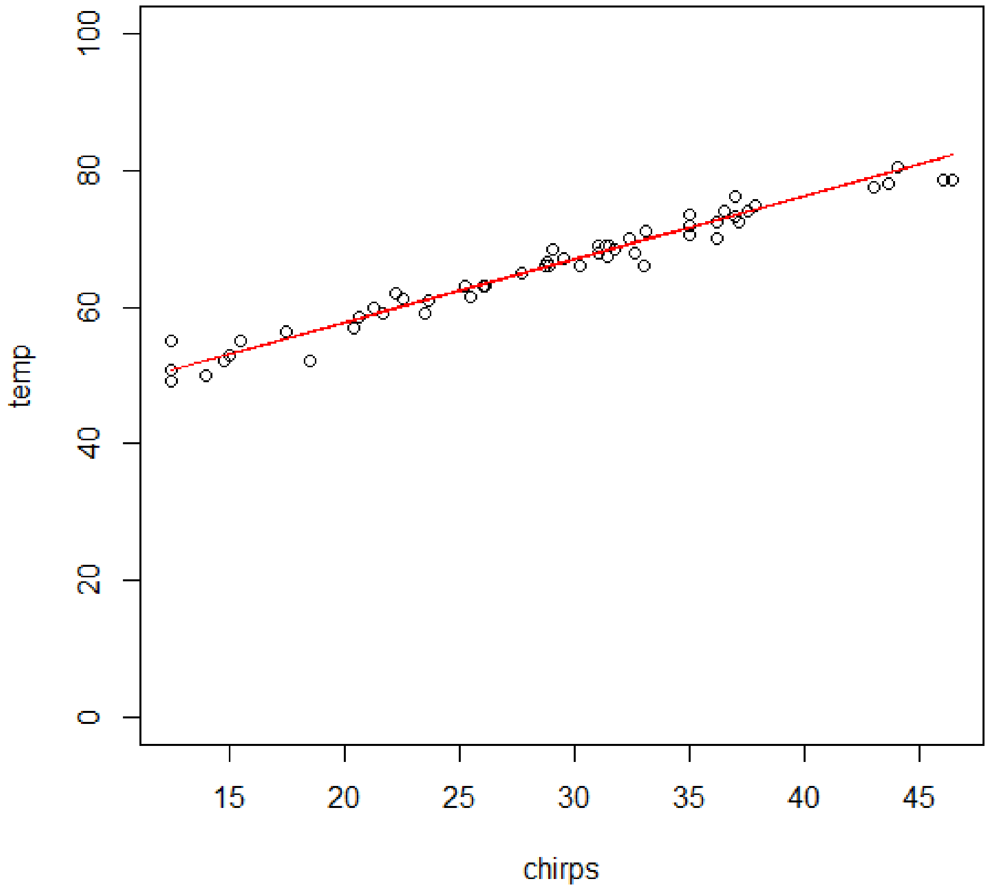
The Solution



Temp = 39.1442 + 0.93Chirps

A Picture

Temp = 39.1442 + 0.93Chirps



How wrong are we???????

R Can Do This

- The command is called rq

```
> rq(temp~chirps)
Call:
rq(formula = temp ~ chirps)

Coefficients:
(Intercept)      chirps
  39.126524    0.929878
```

- Slightly different from what Excel got-the function is not continuous so slight differences when we try to minimize.

Note: R Modeling notation

- R uses the notation y~x1+x2+x2… for all their modeling routines.
- We cover this notation a lot more in Stat 139.
- Can do y~x1*x2 to fit the model
  - Y=b0+b1x1+b2x2+b3x1*x2 and so on.

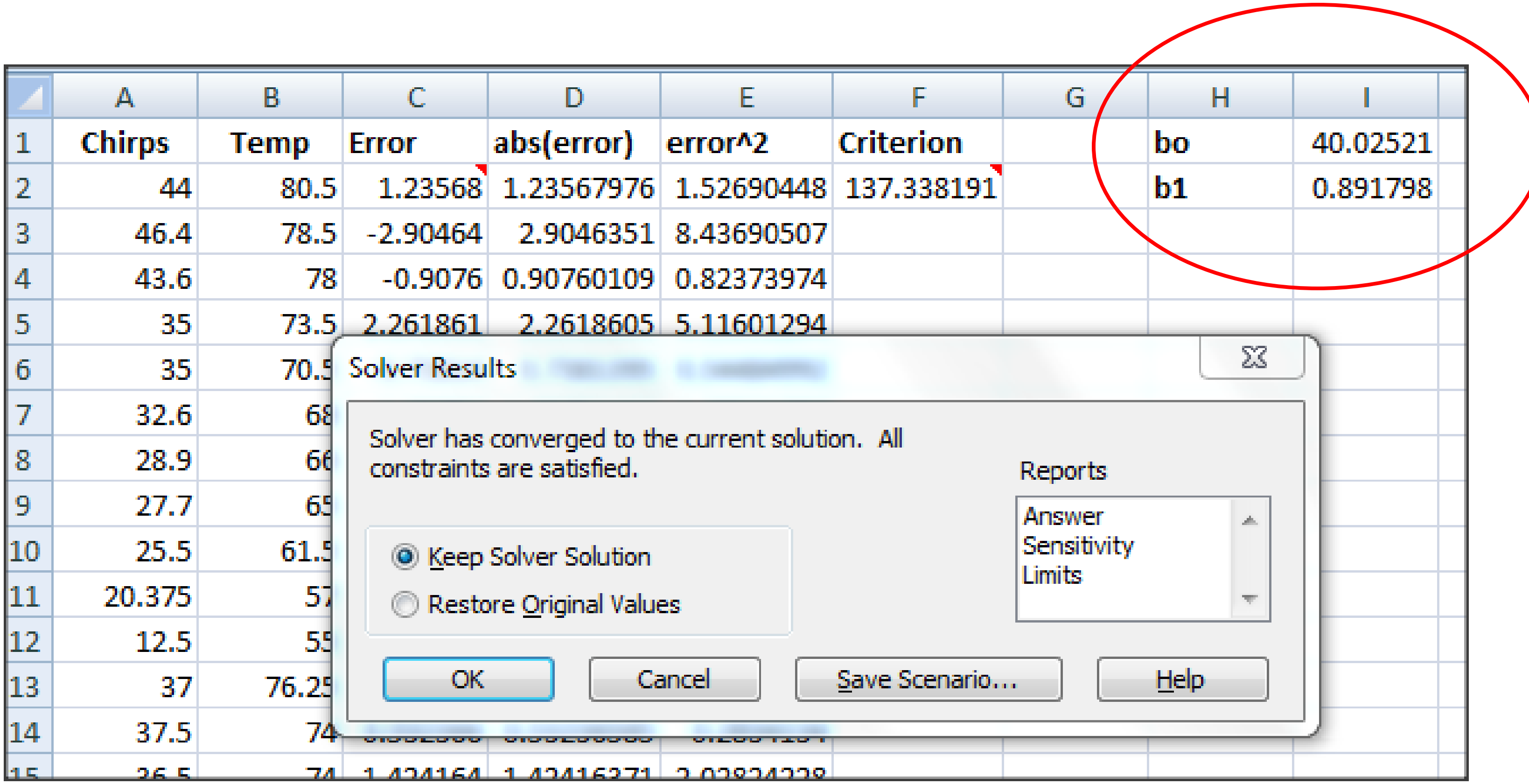
Criterion Function 3 (line method 5)

- The most popular method of fitting a line to data is called the **least-squares method**, and involves solving the following problem

$$\min_{b_0, b_1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

- This can be solved in Excel, or, because it is a continuous criterion function, calculus can be used to find the solution.

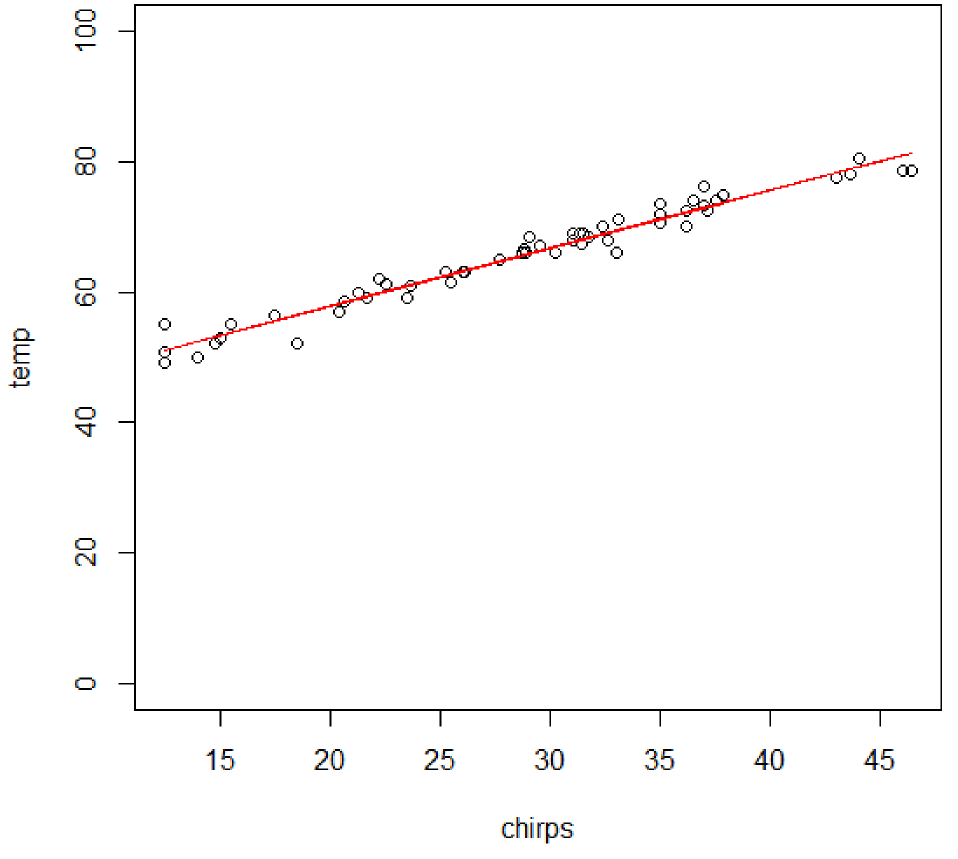
Using Excel



Temp = 40.02 + 0.89Chirps

Nice Picture

Temp = 40.02 + 0.89Chirps



How wrong are we???????

Aside: Using Calculus

- For least squares, we want to solve.

$$\min_{b_0, b_1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

- We know the solution from calculus is to take the first derivative and set it equal to zero.
- This will give us two equations in two unknowns which we can then solve.

The values of b_0 and b_1 which minimize the residual sum of squares are:

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = r\,\frac{s_y}{s_x}$$

Hmm relationship between r and b_1 –does that make sense?

$$b_0 = \bar{Y} - b_1\bar{X}$$

These formulas can be derived using calculus- we pass.

These formulas are the intercept and slope for the “best fitting line”.

Built into R

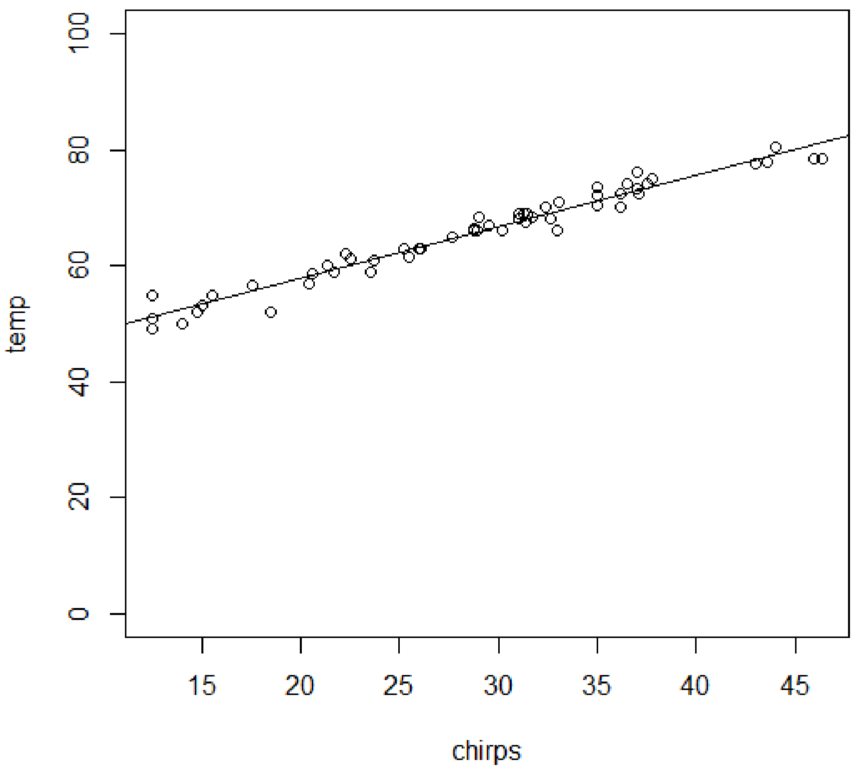
- These equations are built into Stata (and R and many other packages) and are what R uses when you call the lm command

```
> lm(temp~chirps)

Call:
lm(formula = temp ~ chirps)

Coefficients:
(Intercept)       chirps
    40.0252       0.8918
```

Note-these are the same answers Excel gave us using Solver.

R Code for Fitted Line Plot

- For the lm() command, it is very easy to create a fitted line plot

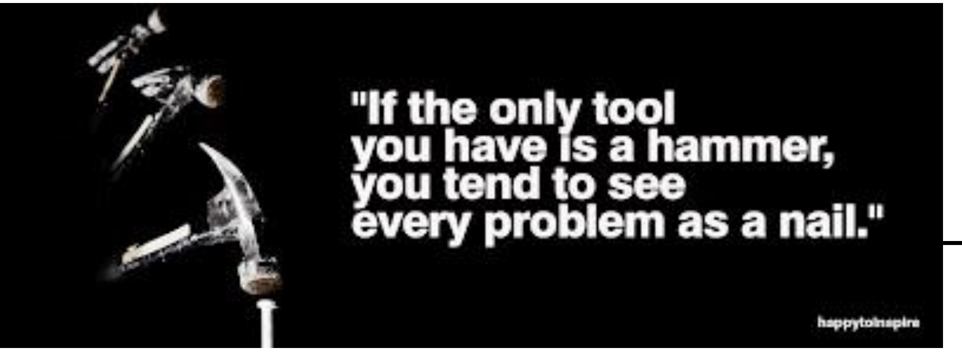


Summary

- Method 1: too stupid to use
- Method 2: temp = 39.8+0.9(chirps)
- Method 3: temp = 29.9+1.24(chirps)[stupid]
- Method 4: temp = 39.14+0.93(chirps)
- Method 5: temp = 40.02+0.89(chirps)
- Why do we care? If we get the same answer different ways-maybe it's a good answer.
- What if there are major differences? Which answer is correct? Hmmm.
- How wrong are we??? (bootstrap…..)

Least Squares and Your Toolbox

- People love least squares-it's the most popular way to fit a line to data, and the method we spend a lot of time examining in this course.
- But it has issues; it used to be the easiest to compute but that's not an issue anymore.
- Always remember



Confidence Intervals

- One reason people like the method of least squares is that there is considerable theory behind the method, including ways to get confidence intervals.
- We will cover the actual theory next time.

```
> fit=lm(temp~chirps)
> confint(fit)
                 2.5 %     97.5 %
(Intercept) 38.5326971 41.5177990
chirps       0.8422354  0.9413595
```

Bootstrap Regression Output

- There are many ways to bootstrap regression output.
- The simplest is to resample the rows of your data.
- That is, define $z_i=\{y_i,x_i\}$ and resample the z_i's.

Bootstrap Code

- Function to return the slope from the lm command:

```
myfunc=function(data,i) {
    x=data[i,1]
    y=data[i,2]
    fit=lm(y~x)
    return(coef(fit)[2])
}
```

Bootstrap Code

```
> mydata=cbind(chirps,temp)
> boots=boot(mydata,myfunc,R=1000)
> boot.ci(boots)
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates

CALL :
boot.ci(boot.out = boots)

Intervals :
Level      Normal              Basic
95%   ( 0.8317,  0.9515 )   ( 0.8323,  0.9494 )

Level     Percentile            BCa
95%   ( 0.8342,  0.9513 )   ( 0.8329,  0.9499 )
```

Regression Example: Market Model

In finance, a popular model is to regress stock returns against returns of some market index, such as the S&P 500.

The slope of the regression line, referred to as "beta", is a measure of how sensitive a stock is to movements in the market.

$$Stockreturn_t = \alpha + \beta Indexreturn_t$$

Market Model

$$Stockreturn_t = \alpha + \beta Indexreturn_t$$

Beta=0 : cash under the mattress

Beta=1 : same risk as the market

0<Beta<1 : safer than the market

Beta >1: riskier than the market

Beta < 0 : what would this mean???

Leveraged ETFs are the RAGE

- FAZ = -3 Financial Index
- FAS = +3 Financial Index
- SDS = -2 S&P500
- SSO = +2 S&P500
- DDM = +2 DJ30
- DXD = -2 DJ30

Fund Summary

The fund seeks daily investment results, before fees and expenses, of 300% of the inverse of price performance of the Russell 1000 Financial Services index. The fund normally creates short positions by investing at least 80% of net assets in financial instruments that, in combination, provide leveraged and unleveraged exposure to the index. It is nondiversified.

They track sort of close…

SYMBOL	TIME & PRICE		CHG & % CHG	
SPY	04:00pm EDT	117.74	↑ 1.07	↑ 0.92%
SSO	04:00pm EDT	40.96	↑ 0.72	↑ 1.78%
SDS	04:00pm EDT	24.11	↓ 0.44	↓ 1.78%

SYMBOL	TIME & PRICE		CHG & % CHG	
DIA	04:00pm EDT	110.92	↑ 0.41	↑ 0.37%
DDM	4:00PM	50.00	↑ 0.38	↑ 0.77%
DXD	4:00PM	20.00	↓ 0.16	↓ 0.79%

ETFs are Very Popular

Vanguard Adds 9 S&P Equity ETFs

etfguide®

Share retweet 3 Email Print

Companies: SPDR S&P 500 | Vanguard 500 Index Investor
Topics: ETFs | Investing Ideas & Strategies

Related Quotes

Symbol	Price	Change
SPY	111.48	+0.56

Symbol	Price	Change
VFINX	102.57	+0.50

ETFguide.com, On Friday September 10, 2010, 12:57 pm EDT

SAN DIEGO (ETFguide.com) - The Vanguard Group launched nine equity ETFs linked to S&P indexes yesterday. The new ETFs follow large, mid and small company U.S. stocks with growth, value and blend characteristics.

The Vanguard S&P 500 ETF (NYSEArca: VOO - News), which also debuted, now serves as the ETF share class for the company's flagship Vanguard 500 Index Fund (Nasdaq: VFINX - News). At the end of August the latter had over $86 billion in assets.

'The new Vanguard index funds and ETFs offer our trademark low costs and tax efficiency, and aim for the utmost tracking precision. They will appeal to financial advisors and institutional investors seeking to build portfolios based on S&P benchmarks. The new ETFs will help Vanguard continue to build momentum in the ETF marketplace,' said Vanguard Chairman and CEO Bill McNabb.

Market Model for SDS and SSO

- SPY is the index, SSO is +2 and SDS is -2 (theoretically).
- Correlation is NOT Beta

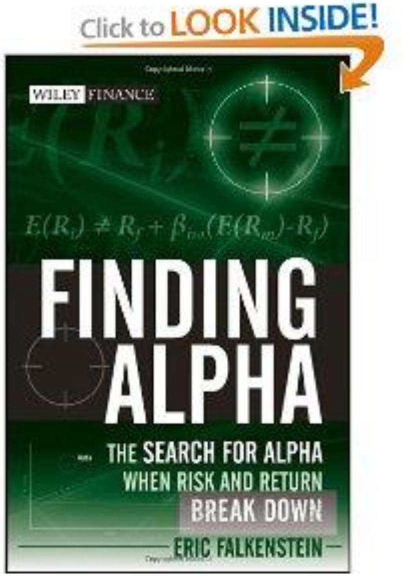
```
. corr
(obs=432)

             |      spy      sds      sso
-------------+---------------------------
         spy |   1.0000
         sds |  -0.9977   1.0000
         sso |   0.9976  -0.9989   1.0000
```

The Search for Alpha

- In the market model, what is the stock return if the index does nothing?

$$Stockreturn_t = \alpha + \beta Indexreturn_t$$

People talk about “buying someone’s alpha”; i.e. what does the fund manager bring to the table above the index returns.



Finding the beta for $MRK

- $MRK is MERCK

Merck & Co. Inc. (MRK) - NYSE Watchlist | Add to Portfolio | Like 181

49.60 0.99 (1.96%) Sep 25, 4:02PM EDT

After Hours : **49.65** 0.05 (0.10%) Sep 25, 7:56PM EDT

Prev Close:	50.59	Day's Range:	49.13 - 51.32
Open:	50.92	52wk Range:	45.69 - 63.62
Bid:	49.63 x 300	Volume:	15,322,722
Ask:	49.75 x 100	Avg Vol (3m):	11,416,600
1y Target Est:	63.20	Market Cap:	139.71B
Beta:	0.580286	P/E (ttm):	14.49
Next Earnings Date:	27-Oct-15	EPS (ttm):	3.42
		Div & Yield:	1.80 (3.56%)

Merck & Company, Inc. Common St

1d 5d 1m 3m 6m 1y 2y 5y max

Quotes delayed, except where indicated otherwise. Currency in USD.

Finding Beta

- Beta is usually calculated using three years of monthly returns

```
> ticker="MRK"
> stock=getSymbols(ticker,from="2012-09-01",auto.assign=FALSE)
> spy=getSymbols("SPY",from="2012-09-01",auto.assign=FALSE)
> stockret=monthlyReturn(Ad(stock))
> spyret = monthlyReturn(Ad(spy))
> n=length(spyret)
> fit=lm(stockret[-n]~spyret[-n])
> coef(fit)
(Intercept)  spyret[-n]
0.003316214 0.572187183
```

There are no Stat Police

- Beta depends on the data you use!

Apple Inc. (AAPL) - NasdaqGS ★ **Watchlist**

114.71 ↓0.29 (0.25%) Sep 25, 4:00

After Hours : **114.72** ↑0.01 (0.01%) Sep 25, 7:59PM ED

Prev Close:	115.00	Day's Ra
Open:	116.36	52wk Ra
Bid:	114.44 x 500	Volume:
Ask:	114.76 x 400	Avg Vol (
1y Target Est:	145.77	Market C
Beta:	1.09091	P/E (ttm)
Earnings Date:	Oct 19 - Oct 23 (Est.)	EPS (ttm
		Div & Yie

114.71 -0.29 (-0.25%)

Sep 25 - Close

NASDAQ real-time data - Disclaimer

Currency in USD

Range	114.02 - 116.69	Div/yield	0.52/1.81
52 week	92.00 - 134.54	EPS	8.66
Open	116.44	Shares	5.70B
Vol / Avg.	56.15M/64.49M	Beta	0.94
Mkt cap	654.16B	Inst. own	60%
P/E	13.25		

From google finance

From yahoo finance

Two Functions

```
gbeta=function(ticker) {
stock=getSymbols(ticker,from="2010-09-01",auto.assign=FALSE)
spy=getSymbols("SPY",from="2010-09-01",auto.assign=FALSE)
stockret=monthlyReturn(Ad(stock))
spyret = monthlyReturn(Ad(SPY))
n=length(spyret)
fit=lm(stockret[-n]~spyret[-n])
cat("The (5 year) Beta for ",ticker," is ",coef(fit)[2],"\n")
##return(coef(fit)[2])
}


ybeta=function(ticker) {
stock=getSymbols(ticker,from="2012-09-01",auto.assign=FALSE)
spy=getSymbols("SPY",from="2012-09-01",auto.assign=FALSE)
stockret=monthlyReturn(Ad(stock))
spyret = monthlyReturn(Ad(spy))
n=length(spyret)
fit=lm(stockret[-n]~spyret[-n])
cat("The (3 year) Beta for ",ticker," is ",coef(fit)[2],"\n")
##return(coef(fit)[2])
}
```

CAPM.beta

- There is a CAPM.beta routine in R
- But all its doing is running a regression behind the scenes
- In package PerformanceAnalytics

```
> CAPM.beta(dailyReturn(SDS),dailyReturn(SPY))
[1] -2.006252
> CAPM.beta(dailyReturn(SSO),dailyReturn(SPY))
[1] 1.992183
```

Also CAPM.beta.bull (bear)

- One could look at “upside beta” versus “downside beta”.
- That is only calculate beta when the index has positive (or negative returns).
- What do you find? [CAPM.beta.bull or bear]

Formally, The Market Model

- We can model the return on a stock (mutual fund, hedge fund, etc…) as

$$R_i = \alpha_i + \beta_i R_M + \varepsilon_i$$

- where

α_i = part of the security's return independent of the market

R_M = rate of return on the market index

β_i = a constant that measures the expected change in R_i given a change in R_M

ε_i = randome noise, indepent of R_M (firm-specific surprises)

Statistical Properties of the MM

- What does the market model say about the variance of returns?
- According to the model

$$
\begin{aligned}
Var(R_i) &= Var(\alpha_i + \beta_i R_M + \varepsilon_i) \\
&= Var(\beta_i R_M + \varepsilon_i) \\
&= \beta_i^2 Var(R_M) + Var(\varepsilon_i) + 2\beta_i Cov(R_M, \varepsilon_i) \\
&= \beta_i^2 Var(R_M) + Var(\varepsilon_i)
\end{aligned}
$$

Add some new notation

- We have

$$Var(R_i) = \beta_i^2 Var(R_M) + Var(\varepsilon_i)$$

- Which we will write as

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma_{\varepsilon,i}^2$$

Decompose the Risk of Security i

- We can decompose the risk (variance) for security *i*:

$$\sigma_i^2 = \underline{\beta_i^2 \sigma_M^2} + \sigma_{\varepsilon,i}^2$$

Part of security i's variance due to market news and movement-**called market** or **systematic risk or non-diversifiable risk.**

Part of security i's variance due to non-market news (e.g. industry effects)-**called nonmarket or unsystematic risk or diversifiable risk.**

Interpretation of Beta

- Examine this formula one last time

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma_{\varepsilon,i}^2$$

- Note that the market variance is the same for all securities.
- So, we say Beta measures a stock's degree of systematic or market risk.

Firm Specific versus Systematic Risks

- **Systematic** (market or macroeconomic) risk usually accounts for about 25% of a typical stock's total risk.
- **Firm-specific** risk usually accounts for about 75% of a typical firm's total risk.
- You can eliminate almost all the firm-specific risk by holding a well-diversified portfolio of stocks. The random bad and good firm-specific events are then largely offset within the portfolio.

Firm Specific versus Systematic Risk

- Most people would never put all of their wealth in one common stock.
  - Too much exposure to firm-specific risk
- However, many people would put a large amount of their wealth in a mutual fund that invests in a diversified portfolio of common stocks.
  - The relevant risk here is the systematic risk, as the firm-specific risk has been diversified away.
- **You can never diversify away systematic risk, as all stocks have exposure to this risk.**

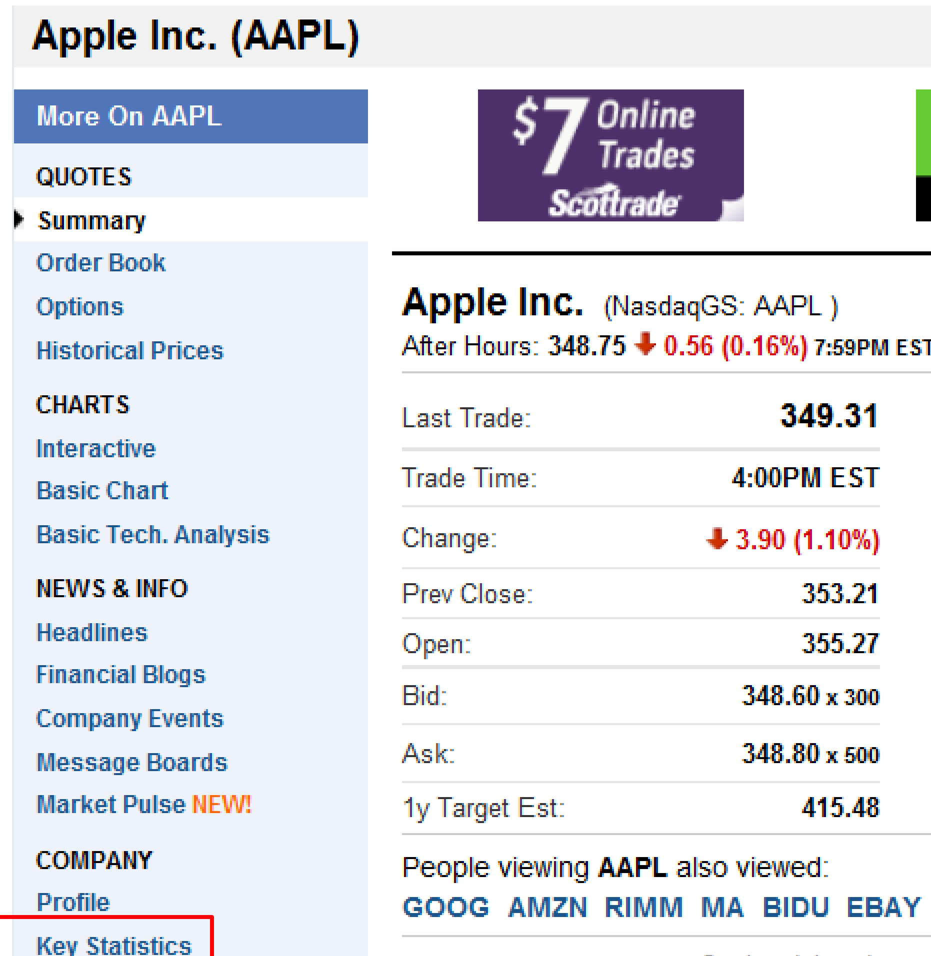
The Beta for AAPL

- Run the regression (older data)

```
> fit=lm(aaplret~spyret)
> fit

Call:
lm(formula = aaplret ~ spyret)

Coefficients:
(Intercept)       spyret
    0.03225      1.20728
```

Compare with Yahoo

- Click on “Key Statistics”



AAPL Beta from Yahoo

Trading Information	
Stock Price History	
Beta:	1.20
52-Week Change[3]:	67.25%
S&P500 52-Week Change[3]:	16.81%
52-Week High (Feb 16, 2011)[3]:	364.90
52-Week Low (May 6, 2010)[3]:	199.25
50-Day Moving Average[3]:	346.42
200-Day Moving Average[3]:	309.23
Share Statistics	

We agree!
But note that Yahoo (and every web site for that matter) DOES NOT give a confidence interval for Beta. Why not?????

A Confidence Interval for Beta

- Its simply beta+/-2(standard error), but R can compute it automatically.

```
> confint(fit)
                  2.5 %     97.5 %
(Intercept) 0.007070099 0.05743848
spyret      0.795459482 1.61909834
```

- What does this imply??

Check on R^2

- According to the regression output, 50% of AAPL's variability is explained by market movements.

```
> summary(fit)

Call:
lm(formula = aaplret ~ spyret)

Residuals:
      Min        1Q    Median        3Q       Max
-0.241825 -0.038749  0.005398  0.044820  0.131521

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.03225    0.01241   2.600   0.0136 *
spyret       1.20728    0.20286   5.951 8.95e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07544 on 35 degrees of freedom
Multiple R-squared: 0.503,     Adjusted R-squared: 0.4888
F-statistic: 35.42 on 1 and 35 DF,  p-value: 8.947e-07
```

Let's find the Beta for JNJ

- JNJ (johnson and johnson), sort of a boring stock.

```
getSymbols("JNJ",from="2008-03-01")
getSymbols("SPY",from="2008-03-01")
spyret=monthlyReturn(SPY)
jnjret=monthlyReturn(JNJ)
```

Run the regression, compare to Yahoo

```
> fit=lm(jnjret~spyret)
> fit

Call:
lm(formula = jnjret ~ spyret)

Coefficients:
(Intercept)       spyret
 -0.0002453    0.5776404
```

Yup, we got this beta thing down I think

Trading Information	
Stock Price History	
Beta:	0.58
52-Week Change[3]:	-4.32%
S&P500 52-Week Change[3]:	16.81%
52-Week High (Apr 20, 2010)[3]:	66.20
52-Week Low (Jul 22, 2010)[3]:	56.86
50-Day Moving Average[3]:	61.17

Always give a confidence interval

- Estimates by themselves are always useless!
- Always report a confidence interval

```
> confint(fit)
                  2.5 %     97.5 %
(Intercept) -0.01136636 0.01087579
spyret       0.39578529 0.75949558
> 
```

Check on R^2

- Based on the regression output below, 54% of the variation of JNJ's returns are related to movements in the market.

```
> summary(fit)

Call:
lm(formula = jnjret ~ spyret)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.070986 -0.018944  0.006906  0.019956  0.069626 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.0002453  0.0054781  -0.045    0.965    
spyret       0.5776404  0.0895791   6.448 1.99e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03331 on 35 degrees of freedom
Multiple R-squared: 0.543,     Adjusted R-squared: 0.5299 
F-statistic: 41.58 on 1 and 35 DF,  p-value: 1.991e-07
```

Portfolio Betas

- Consider an equally weighted portfolio of AAPL, JNJ, SBUX and MSFT.

```
> portret = (aaplret+jnjret+sbuxret+msftret)/4
> portbeta=coef(lm(portret~spyret))[2]
> portbeta
  spyret
1.005001
```

- It turns out that Betas add.

Betas Add

```
> aaplbeta=coef(lm(aaplret~spyret))[2]
> jnjbeta=coef(lm(jnjret~spyret))[2]
> msftbeta=coef(lm(msftret~spyret))[2]
> sbuxbeta=coef(lm(sbuxret~spyret))[2]
> (aaplbeta+jnjbeta+msftbeta+sbuxbeta)/4
  spyret
1.005001
> portbeta
  spyret
1.005001
```

Why Least Squares?

- As a quick aside, (least squares) regression fits a line by solving the equation

$$\min_{b_0, b_1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

```
lm(y~x)
```

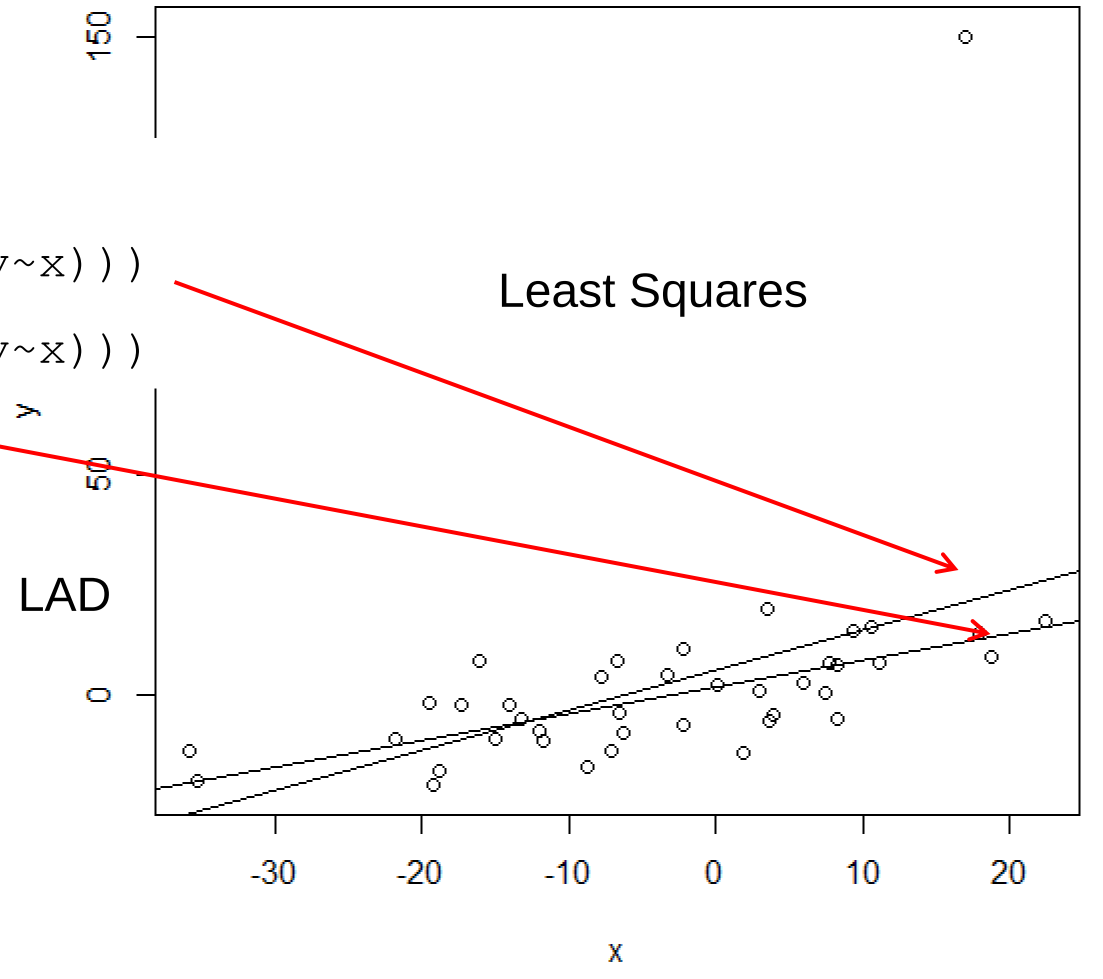
- Why not solve the following equation? This is called LAD (least absolute deviation) regression.

$$\min_{b_0, b_1} \sum_{i=1}^{n} | Y_i - b_0 - b_1 X_i |$$

```
rq(y~x)
[in package
quantreg]
```

Least Squares versus MAD

```
plot(x,y)
lines(abline(lm(y~x)))
lines(abline(rq(y~x)))
```

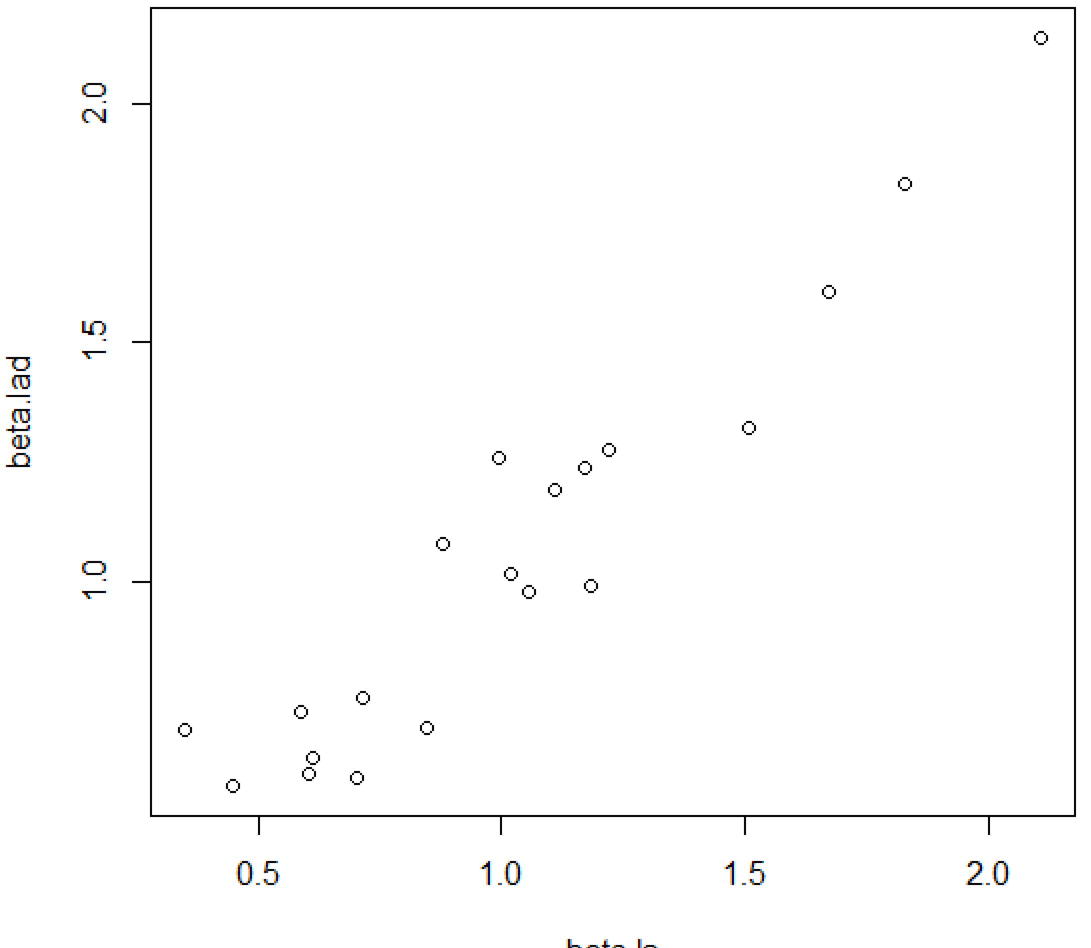


LAD Betas

```
asset.names=c("ATVI", "ADBE", "AKAM", "ALTR", "AMZN",
"AMGN", "APOL", "AAPL", "AMAT", "ADSK", "ADP" , "BIDU",
"BBBY", "BIIB","BMC",  "BRCM" ,"CHRW" ,"CA" ,  "CELG",
"ANF")

n=length(asset.names)
getSymbols("SPY")
spy.ret=monthlyReturn(Ad(SPY))
beta.ls=1:n
beta.lad=1:n

for(i in 1:n){
      x=getSymbols(asset.names[i],auto.assign=FALSE)
       x.ret=monthlyReturn(Ad(x))
      beta.ls[i]=coef(lm(x.ret~spy.ret))[2]
      beta.lad[i]=coef(rq(x.ret~spy.ret))[2]
}
```

Which ones are correct?

```
> cor(beta.ls,beta.lad)
[1] 0.9531647
```



Beta is Beta until its not

- People love talking and thinking about Beta; it reduces the complexities of risk and comparing stocks to a single, easily interpretable number.
- However, Beta has numerous issues.
- Beta depends on the day of the week, the time scale, the market index being used, etc….

Some Issues with Beta

- Which market index?
- Which time intervals?
- Time length of data?
- Non-stationary
  - Beta estimates of a company change over time.
  - How useful is the beta you estimate now for thinking about the future?

The Beta for AAPL

YAHOO

Trading Information	
Stock Price History	
Beta:	1.18
52-Week Change[3]:	50.97%
S&P500 52-Week Change[3]:	11.37%
52-Week High (Feb 16, 2011)[3]:	364.90
52-Week Low (May 6, 2010)[3]:	199.25
50-Day Moving Average[3]:	349.39
200-Day Moving Average[3]:	319.52

Apple Inc. (NasdaqGS: AAPL)

After Hours: **341.00** 0.20 (0.06%) **5:44PM EDT**

Last Trade:	**341.20**	Day's Range:	**339.14 - 342.62**
Trade Time:	**4:00PM EDT**	52wk Range:	**199.25 - 364.90**
Change:	1.90 (0.56%)	Volume:	**11,639,820**
Prev Close:	**339.30**	Avg Vol (3m):	**17,462,300**
Open:	**342.58**	Market Cap:	**314.34B**
Bid:	**341.01 x 100**	P/E (ttm):	**19.04**
Ask:	**341.20 x 1000**	EPS (ttm):	**17.92**
1y Target Est:	**422.53**	Div & Yield:	**N/A (N/A)**

GOOGLE

Apple Inc. (Public, NASDAQ:AAPL) Watch this stock

341.20

+1.90 (0.56%)

Mar 22 - Close

NASDAQ real-time data - Disclaimer

Currency in USD

Range	339.14 - 342.62	P/E	19.05
52 week	199.25 - 364.90	Div/yield	-
Open	342.56	EPS	17.91
Vol / Avg.	11.64M/18.69M	Shares	921.28M
Mkt cap	314.34B	Beta	1.37

Reuters.com just confuses me

VALUATION RATIOS

	Company	Industry	Sector	S&P 500
P/E Ratio (TTM)	18.46	10.85	12.80	17.31
P/E High - Last 5 Yrs.	39.05	29.30	60.30	88.93
P/E Low - Last 5 Yrs.	18.91	10.50	11.61	12.27
Beta	1.39	1.18	1.04	1.32

Wtf? Maybe the average beta of all SP500 stocks??

Let's try GOOG

Trading Information	
Stock Price History	
Beta:	1.04
52-Week Change[3]:	5.01%
S&P500 52-Week Change[3]:	11.37%
52-Week High (Jan 19, 2011)[3]:	642.96
52-Week Low (Jul 6, 2010)[3]:	433.63
50-Day Moving Average[3]:	601.72
200-Day Moving Average[3]:	583.81

VALUATION RATIOS

	Company	Industry	Sector	S&P 500
P/E Ratio (TTM)	21.33	17.55	12.80	17.31
P/E High - Last 5 Yrs.	52.01	45.41	60.30	88.93
P/E Low - Last 5 Yrs.	22.58	12.55	11.61	12.27
Beta	1.20	0.81	1.04	1.32

Google Inc. (Public, NASDAQ:GOOG) Watch this stock

577.32

+0.82 (0.14%)

After Hours: **577.32** 0.00 (0.00%)
Mar 22, 6:19PM EDT
NASDAQ real-time data - Disclaimer
Currency in USD

Range	572.51 - 579.23	P/E	21.95
52 week	433.63 - 642.96	Div/yield	-
Open	577.27	EPS	26.30
Vol / Avg.	1.89M/2.85M	Shares	321.52M
Mkt cap	185.62B	Beta	1.19

What about JNJ?

Stock Price History	
Beta:	0.58
52-Week Change[3]:	-9.63%
S&P500 52-Week Change[3]:	11.37%
52-Week High (Apr 20, 2010)[3]:	66.20
52-Week Low (Jul 22, 2010)[3]:	56.86
50-Day Moving Average[3]:	60.24
200-Day Moving Average[3]:	61.92

VALUATION RATIOS

	Company	Industry	Sector	S&P 500
P/E Ratio (TTM)	12.26	21.69	45.51	17.31
P/E High - Last 5 Yrs.	18.55	27.43	58.69	88.93
P/E Low - Last 5 Yrs.	12.82	11.55	13.57	12.27
Beta	0.59	0.63	0.63	1.32

Johnson & Johnson (Public, NYSE:JNJ) Watch this stock

58.79

-0.04 (-0.07%)

After Hours: **58.81** +0.02 (0.03%)
Mar 22, 6:34PM EDT
NYSE real-time data - Disclaimer
Currency in USD

Range	58.70 - 59.15	P/E	12.31
52 week	56.86 - 66.20	Div/yield	0.54/3.67
Open	58.80	EPS	4.78
Vol / Avg.	10.16M/12.13M	Shares	2.74B
Mkt cap	161.10B	Beta	0.58

Beta's are Weird

- Consider the paper, "Are calculated betas worth for anything?"
- Great reading, particularly since Betas are ubiquitous and easily come up in finance discussions.

Table 1 from the paper

- Consider the following table, showing Betas collected from different web sites:

Table 1. Betas of different companies according to different sources

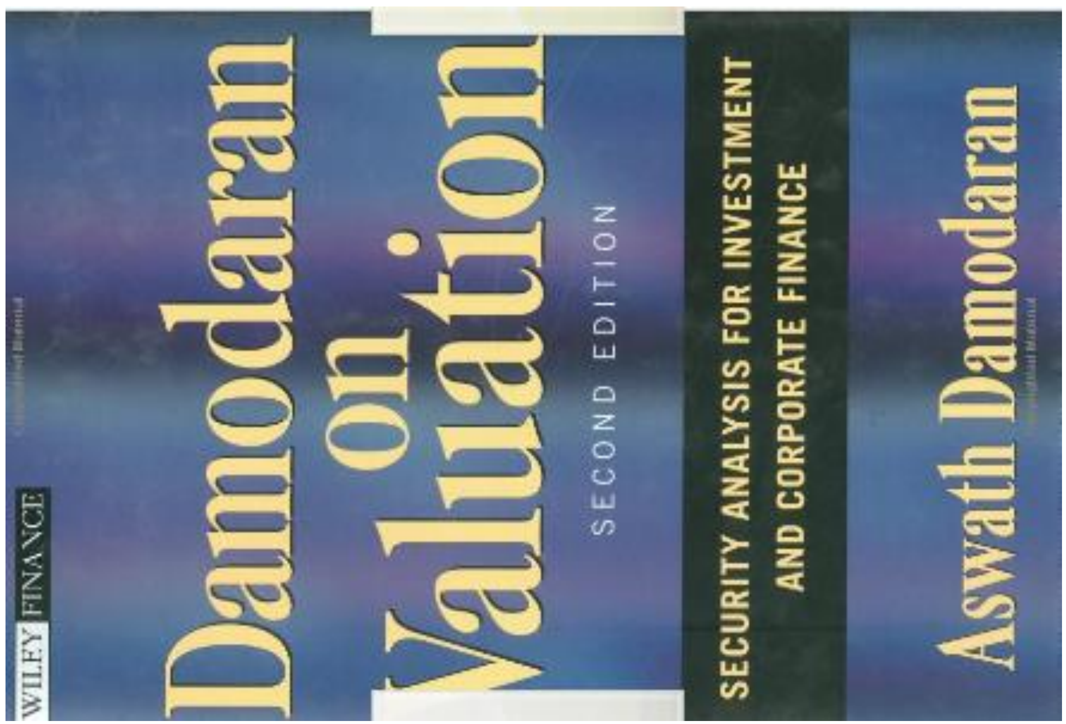
	AT&T	Boeing	CocaCola	Date
Yahoo	0.61	0.46	0.29	12-febr-03
Multex	0.87	0.66	0.42	12-febr-03
Quicken	1.14	0.66	0.41	12-febr-03
Reuters	0.87	0.68	0.42	12-febr-03
Bloomberg	1.00	1.07	0.64	12-febr-03
Datastream	1.10	1.10	0.37	12-febr-03
Buy&hold	0.84	0.66	0.41	14-febr-03

- Which one is correct?

Damodaran, 2001

Beta estimates for Cisco versus the S&P 500. *Source: Damodaran (2001, page 72)*

	Daily	Weekly	Monthly	Quarterly
2 years	1.72	1.74	1.82	2.7
5 years	1.63	1.70	1.45	1.78



Bruner (1998)

In addition to relying on historical data, use of this equation to estimate beta requires a number of practical compromises, each of which can materially affect the results. For instance, increasing the number of time periods used in the estimation may improve the statistical reliability of the estimate but risks the inclusion of stale, irrelevant information. Similarly, shortening the observation period from monthly to weekly, or even daily, increases the size of the sample but may yield observations that are not normally distributed and may introduce unwanted random noise. A third compromise involves choice of the market index. Theory dictates that R_m is the return on the market portfolio, an unobservable portfolio consisting of *all* risky assets, including human capital and other nontraded assets, in proportion to their importance in world wealth. Beta providers use a variety of stock market indices as proxies for the market portfolio on the argument that stock markets trade claims on a sufficiently wide array of assets to be adequate surrogates for the unobservable market portfolio.

Bruner (1998)

Exhibit 4. Compromises Underlying Beta Estimates and Their Effect on Estimated Betas of Sample Companies

	Bloomberg[a]	Value Line	Standard & Poor's
Number	102	260	60
Time Interval	wkly (2 yrs.)	wkly (5 yrs.)	mthly(5 yrs.)
Market Index Proxy	S&P 500	NYSE composite	S&P 500
Mean Beta	1.03	1.24	1.18
Median Beta	1.00	1.20	1.21

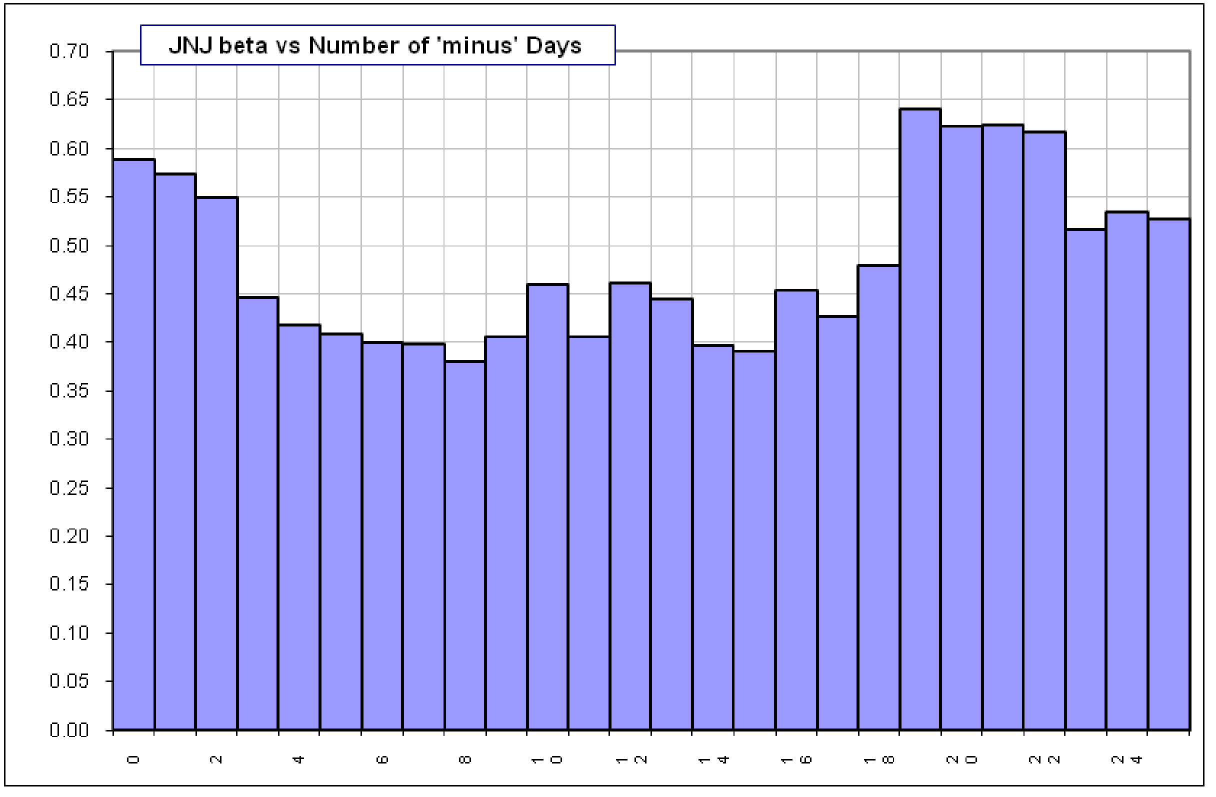
[a]With the Bloomberg service, it is possible to estimate a beta over many differing time periods, market indices, and as smoothed or unadjusted. The figures presented here represent the base-line or default-estimation approach used if other approaches are not specified.

Calculated Betas Change a lot

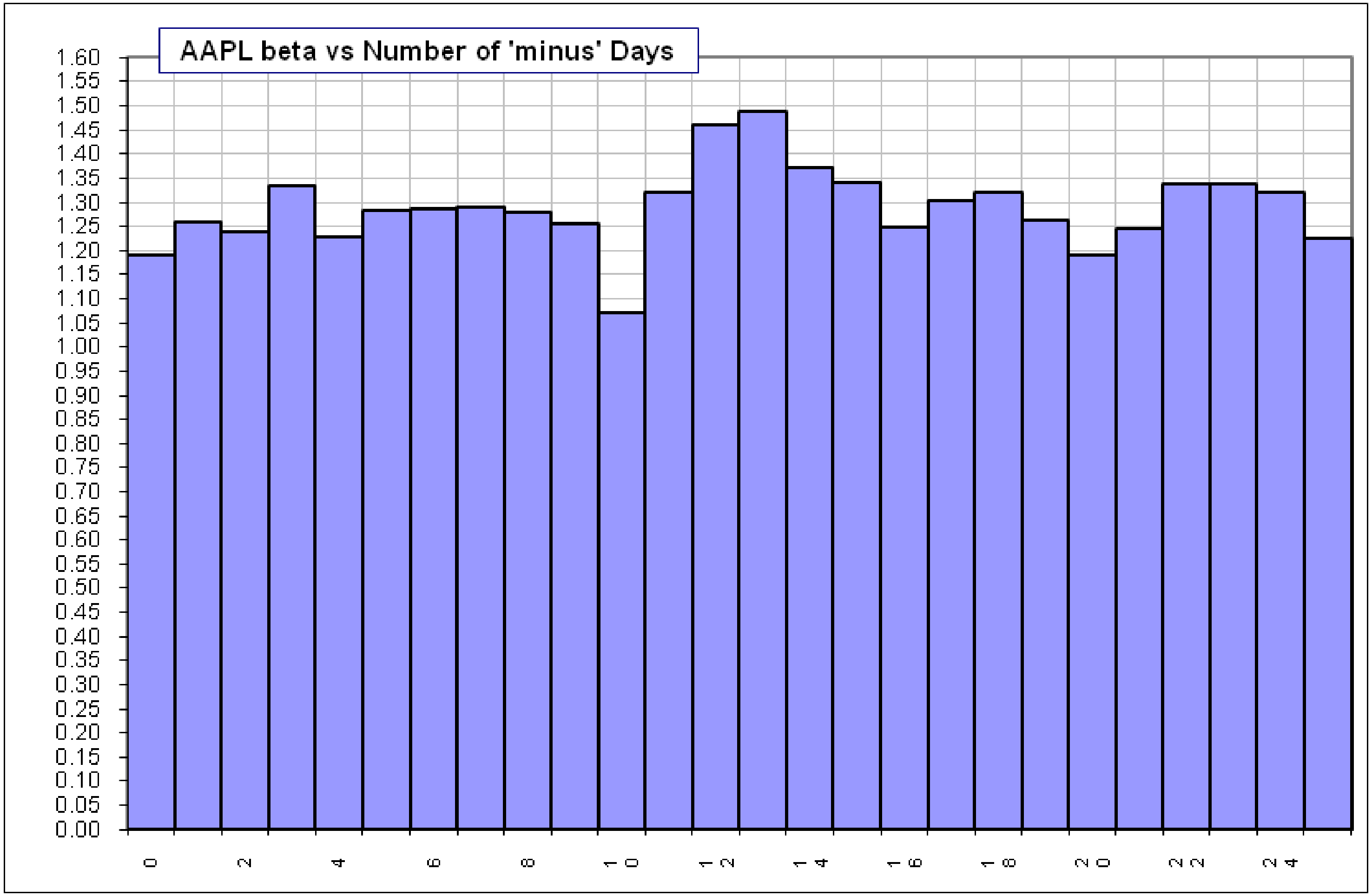
- Gummy and independently Fernandaz had an interesting idea.
- Suppose we are interested in calculating betas on monthly returns.
- The standard is to calculate it based on end of month data (last day of month).
- What if we use the 12^{th} of each month, or 22^{nd}, or something else. Will beta change?

Betas change from 1 day to the next

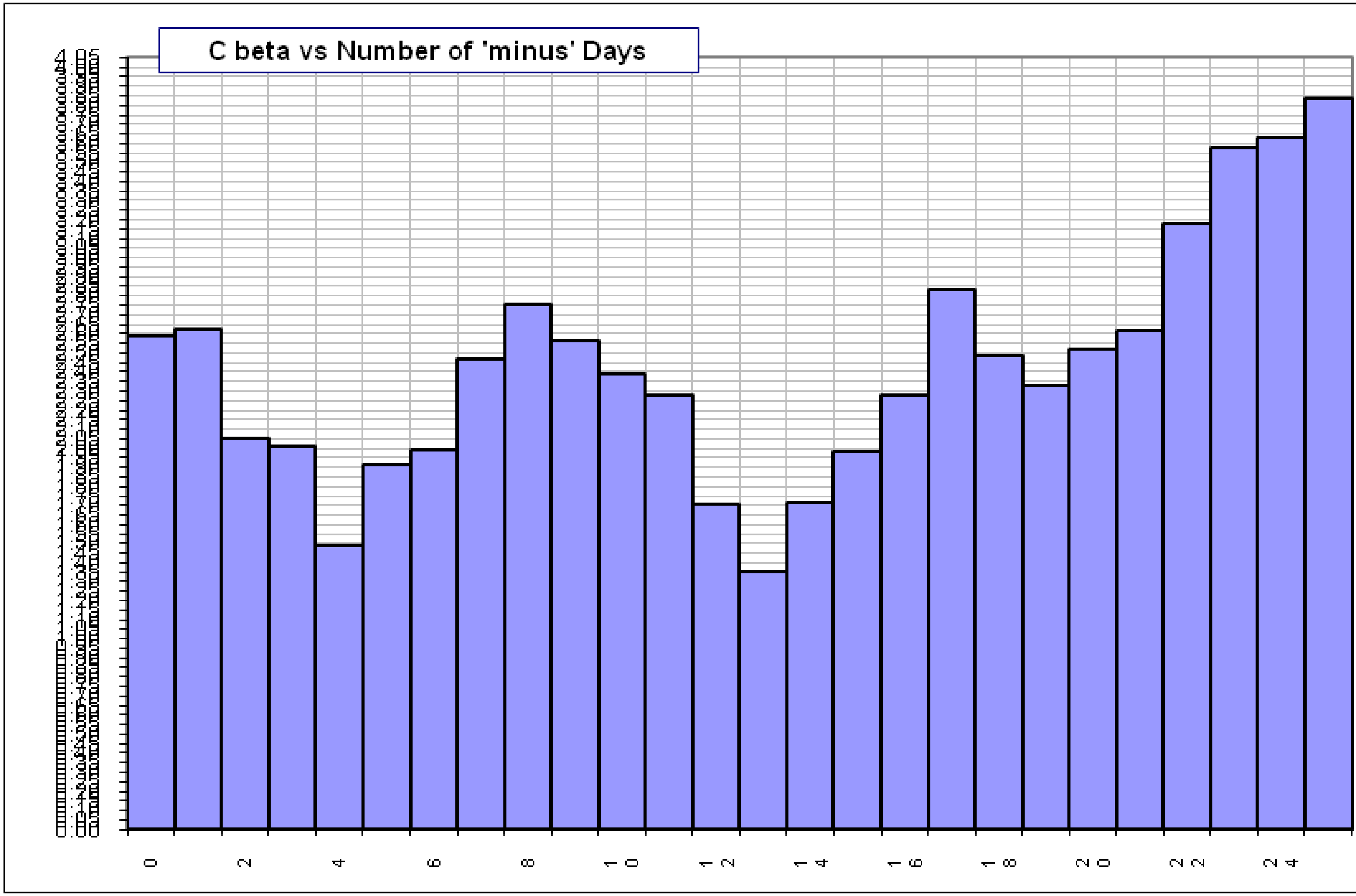
- Gummy has a great spreadsheet that calculates this **weird** result



AAPL Doesn't Change Too Much



But Check out Citibank



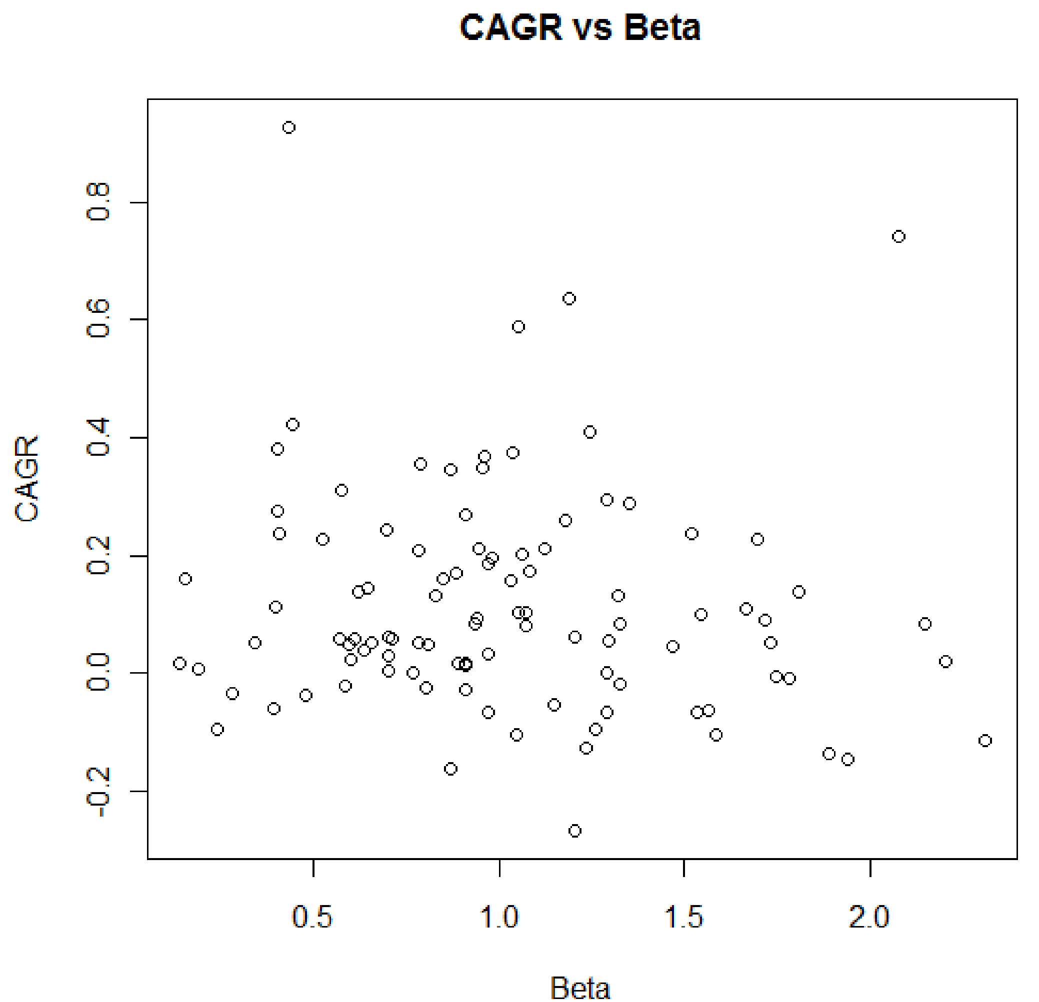
Conclusion

- Calculated Betas are very unstable!

CAGR and Beta

- According to the CAPM, a bigger Beta implies a bigger expected stock return.
- So what should a graph of CAGR (compounded annual return) versus Beta look like?
- Probably not what you expect.

Beta vs CAGR for Nasdaq 100 stocks



To be fair, we should see what this looks like during a bull market like 1996-2000, but it will look even worse.

Conclusion

- Calculated Betas are very unstable!
- So can’t really trust the market model
- What is one to do? Read the appendix in “Are calculated betas worth for anything?”; it gives a very thorough analysis of the research on Beta through the years.

Adjusted Betas

- Betas are non-stationary.
- Researchers have shown that adjusted beta is often a better forecast of future beta than is historical beta. As a consequence, practitioners often use adjusted beta.

Beta as an Autoregressive Process

- One simple idea is to model beta as a first order autoregressive process:

$$\beta_{i,t+1} = \alpha_0 + \alpha_1 \beta_{i,t} + \epsilon_{i,t+1}$$

- It has been academically shown that adjusted beats predict future betas better than historical data since betas are on average mean reverting.

A very simple adjustment

- An adjustment that Merril Lynch and Bloomberg use is as follows.
- Adjusted Beta = 0.333+0.667*hist Beta
- Historical beta=1, adj beta =1
- Historical beta=1.5, adj beta = 1.33
- Historical beta=0.5, adj beta=0.667

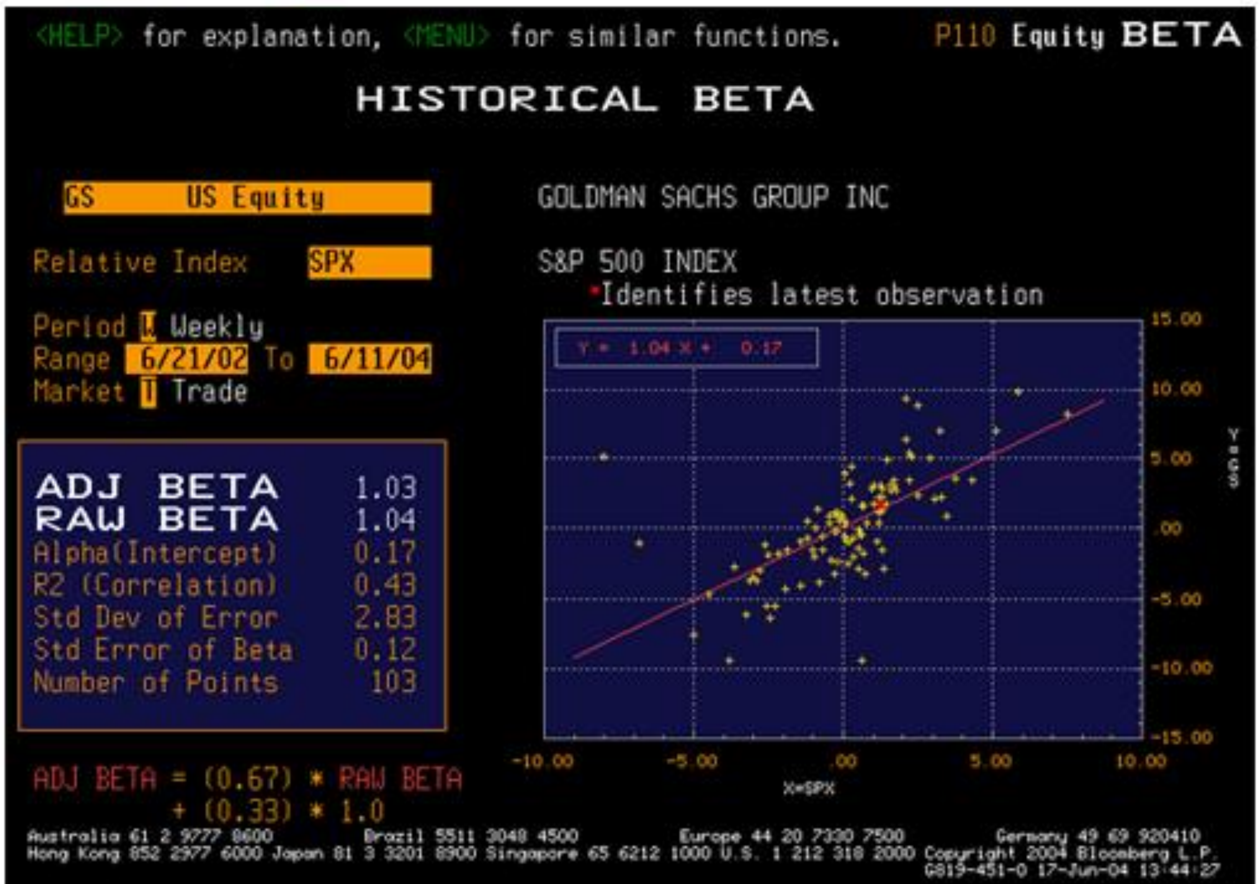
From Bloomberg

Enter this command to look up the beta of a stock:

{ticker symbol} EQUITY **BETA** GO

Example: The following screenshot shows the result for the beta of Goldman Sachs.

GS EQUITY **BETA** GO



Some Silliness

- Bloomberg adj beta = 0.333+0.667*beta
- Value Line adj beta = 0.35 + 0.67*beta
- Merril Lynch adj beta = 0.33743+0.66257*beta
- And there are many more methods out there…
- It is highly believed though that some adjustment of beta should be used instead of using the historical value directly.
- These adjusted betas are then used as before to estimate variances and covariances.