Simple Linear Regression: A Model for the Mean

NOTATION

LEAST SQUARES PRINCIPLE

Terminology & Goals

- There is a lot of notation and vocabulary involved in linear regression
- The core goal behind simple linear regression is estimate a relationship between
  - an input known as the **EXPLANATORY VARIABLE**
  - and another measurement known as the **RESPONSE VARIABLE**
- **Etymology:**
  - **Linear:** We model this relationship as linear for simplicity and interpretability. We must check this modeling assumption.
  - **Regression:** Charles Darwin's cousin, Francis Galton, studied heritability of traits. He found that extra tall people tend to have less tall offspring and extra short people tend to have less short offspring
    $\rightarrow$ Regression

Notation for the Mean

- Y is the response variable
- X is the explanatory variable
- $\mu\{Y|X\}$ is the “mean of Y as a function of X”

For Simple Linear Regression (SLR), we write this mean as

$\mu\{Y|X\} = \beta_0 + \beta_1 X$

- β_0 has the same **units** as Y

(this is the intercept)

- β_1 has the same **units** as Y/X

(this is a rate or slope)

Example: Y (deaths per million) is mortality from skin cancer in a state & X is state latitude (in degrees)

β_0 is in deaths per million
β_1 is in (deaths per million)/degrees

Notation for the Mean

- Y is the response variable
- X is the explanatory variable
- $\mu\{Y|X\}$ is the "mean of Y as a function of X"

Example: Suppose X takes on values $X = \text{“}control\text{”}$ or $X = \text{“}treatment\text{”}$

Re-define: $X = \begin{cases} 0 \; if \; X = \text{“}control\text{”} \\ 1 \; if \; X = \text{“}treatment\text{”} \end{cases}$

$\mu\{Y|X\} = \beta_0 + \beta_1 X$ implies:

- $\mu_{control} = \beta_0$
- $\mu_{treatment} = \beta_0 + \beta_1$

(Generally, it is better to analyze nominal X as an ANOVA and interval/ratio X as a regression. Ordinal X can be analyzed with either, so careful thought is required)

Example: Temperature

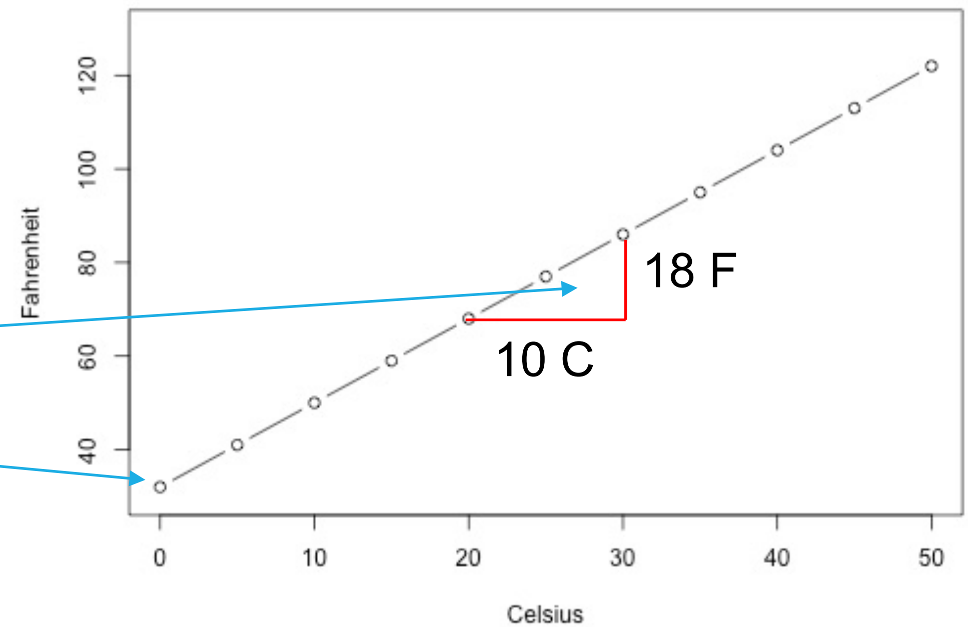
In North America, we have two standards for measuring temperature:

- Fahrenheit (F)
- Celsius (C)

They are related via

the linear relationship

$F = 32 + {}^{9}/_{5}\,\mathrm{C}$



This relationship is exact,

so no statistics is necessary

Notation for the Standard Deviation

Just like $\mu\{Y|X\}$ is the “mean of Y as a function of X”

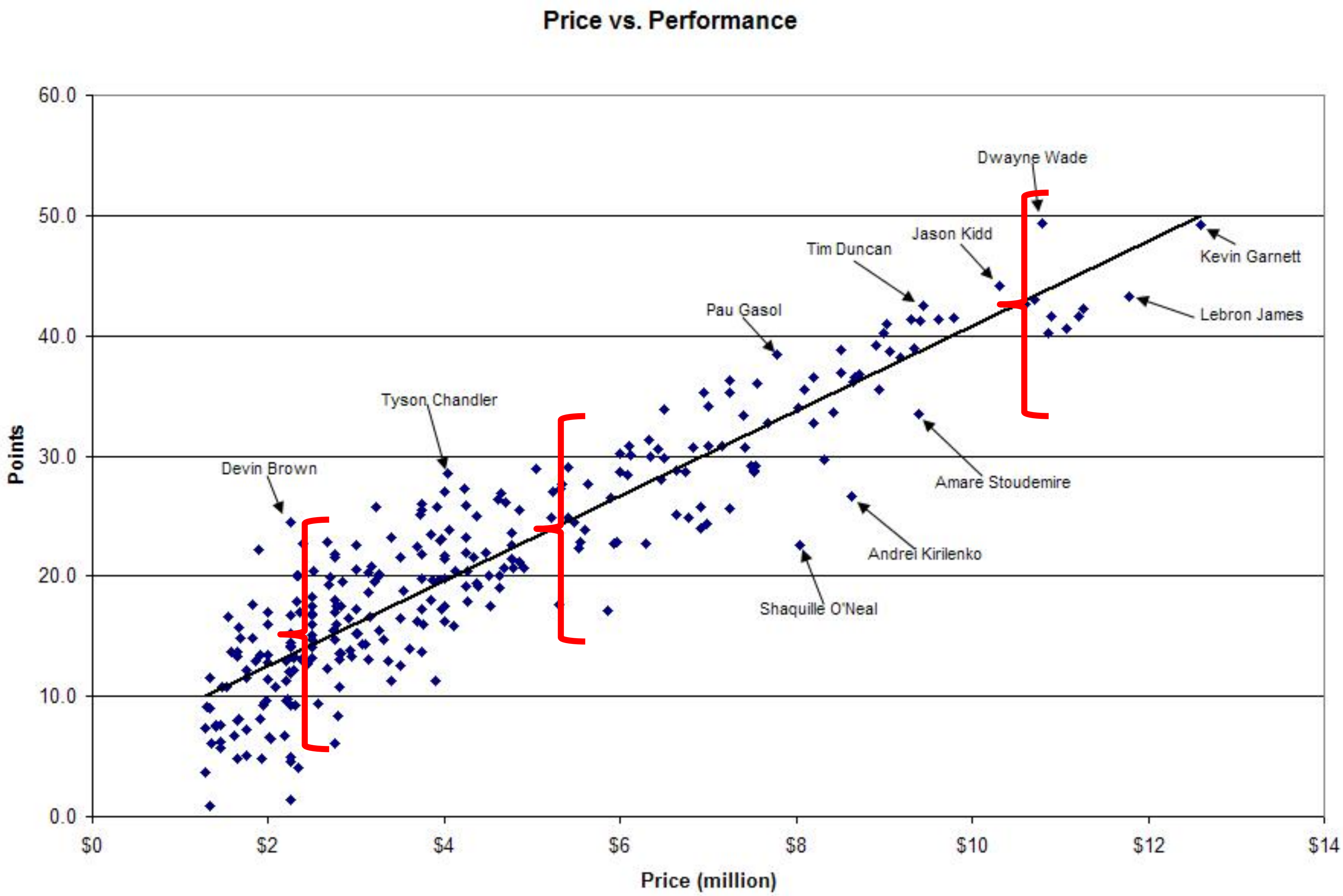
We also have $\sigma\{Y|X\}$ as the standard deviation of Y at X

Reminder: In ANOVA, we assumed that all groups had the same standard deviation (that is, σ doesn’t depend on the value of X)

For Simple Linear Regression, we make the same assumption: $\sigma\{Y|X\} = \sigma$

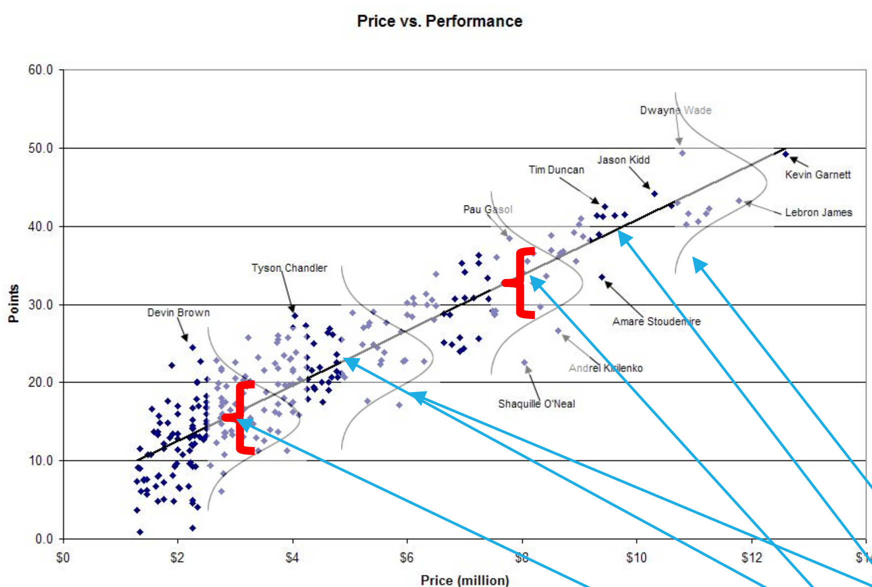
(That is, the standard deviation doesn’t depend on X)

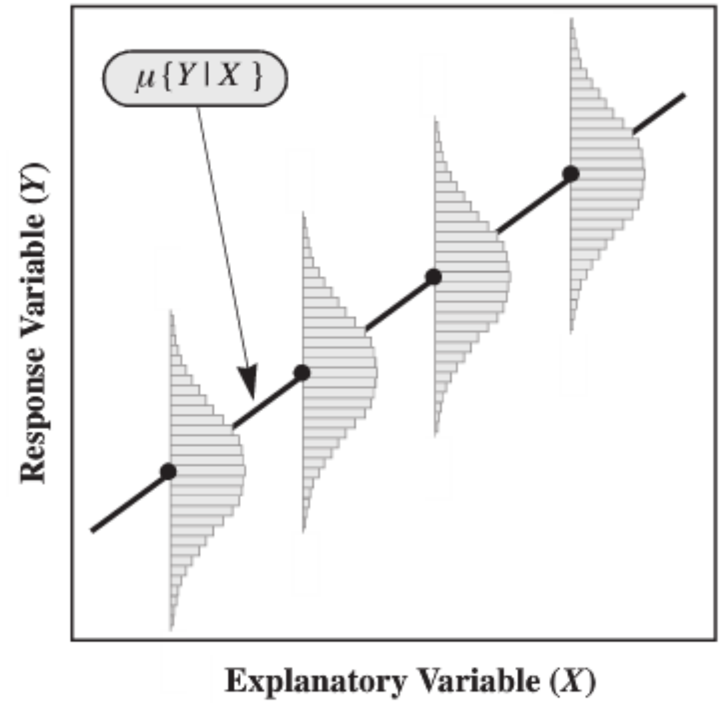
In the temperature example, σ = 0 due to the relationship being exact



Price vs. Performance
Points
60.0
50.0
40.0
30.0
20.0
10.0
0.0
$0
$2
$4
$6
$8
$10
$12
$14
Price (million)
Dwayne Wade
Jason Kidd
Tim Duncan
Kevin Garnett
Lebron James
Pau Gasol
Tyson Chandler
Devin Brown
Amare Stoudemire
Andrei Kirilenko
Shaquille O'Neal

Assumptions



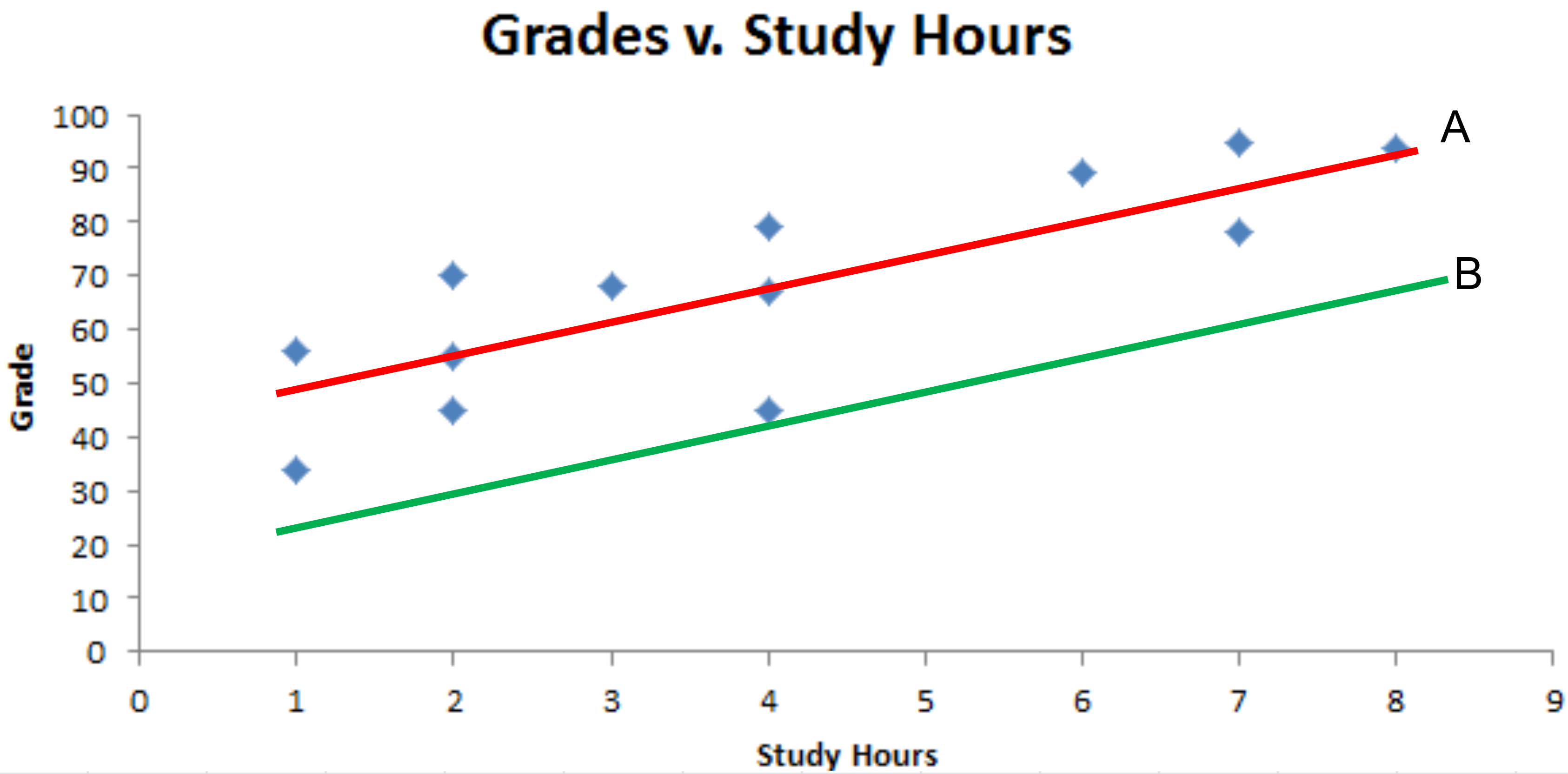


Model Assumptions

1. There is a normally distributed subpopulation of responses for each value of the explanatory variable.
2. The means of the subpopulations fall on a straight line function of the explanatory variable.
3. The subpopulation standard deviations are all equal (to σ).
4. The selection of an observation from any of the subpopulations is independent of the selection of any other observation.

There is an NBA salary cap & only so many points can be scored in a game…

How Do We Estimate the Mean?

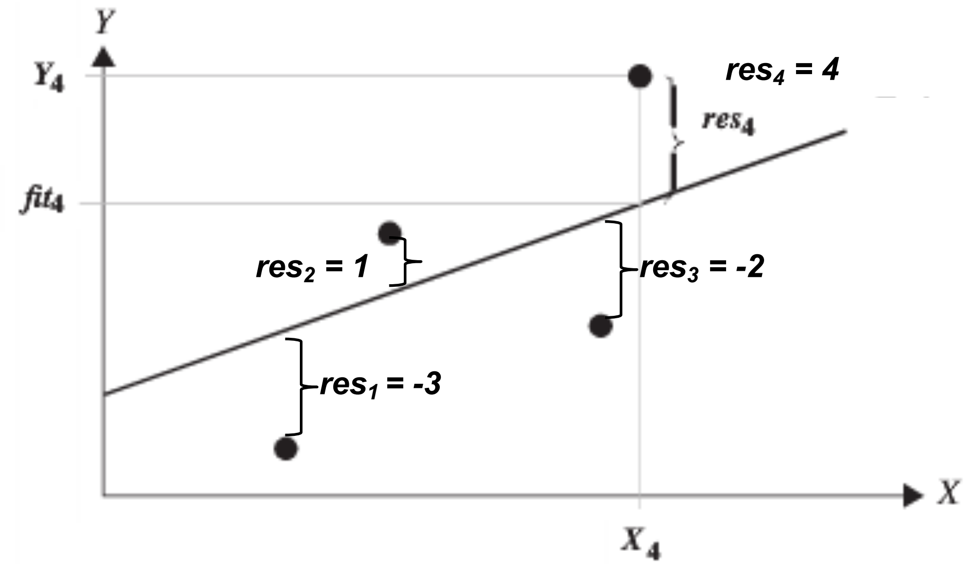


Study Hours	1	1	2	2	2	3	4	4	4	6	7	7	8
Grades	34	56	45	70	55	68	67	79	45	89	95	78	94

How Do We Estimate the Mean?

Remember: Residuals are the difference between an estimate and the actual value

The i^{th} RESIDUAL is defined to be res_i

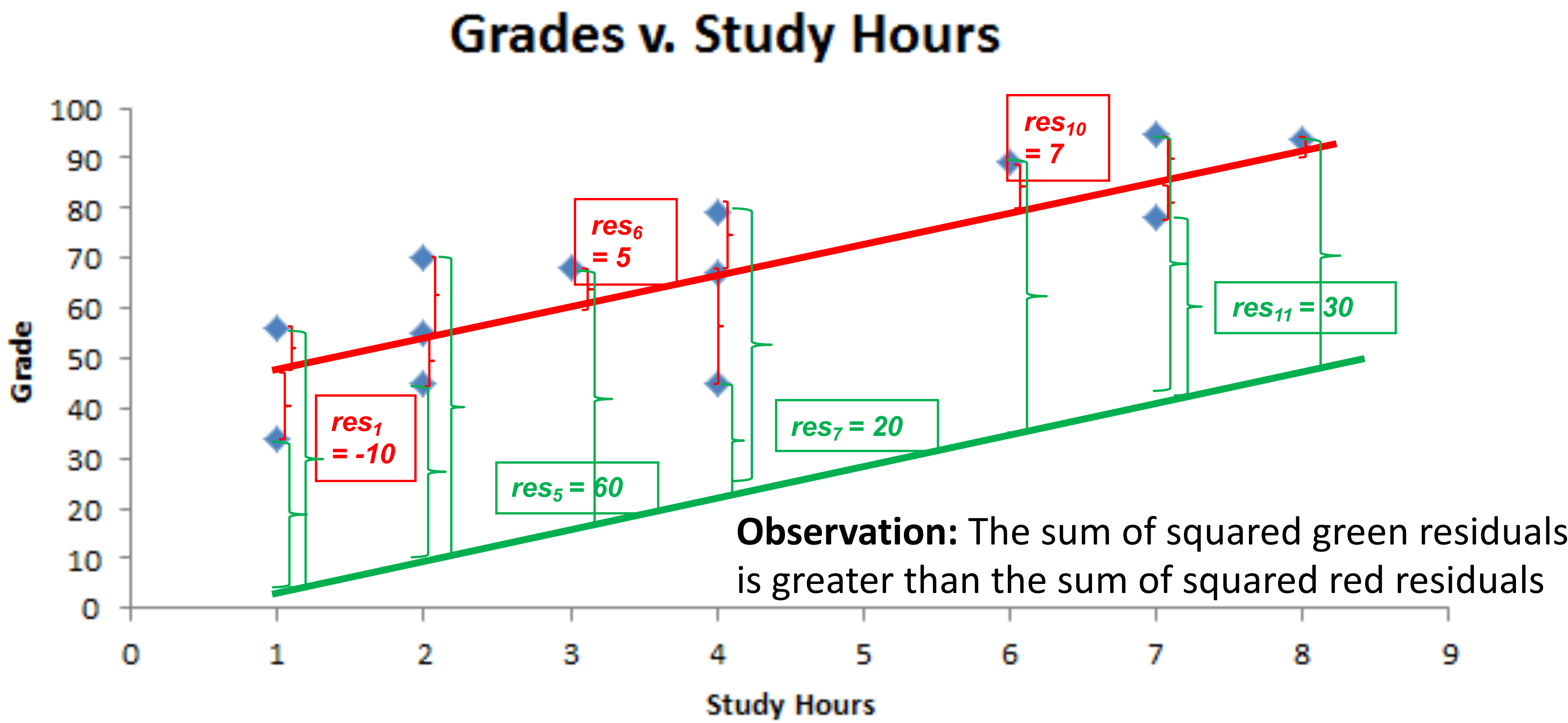


If we simply add the residuals we will get zero
→ square each residual and add them together.

Sum of Squared Residuals = SSR =

$$\sum_{i=1}^{n} res_i^2 = (-3)^2 + (1)^2 + (-2)^2 + (4)^2 = 9 + 1 + 4 + 16 = 30$$

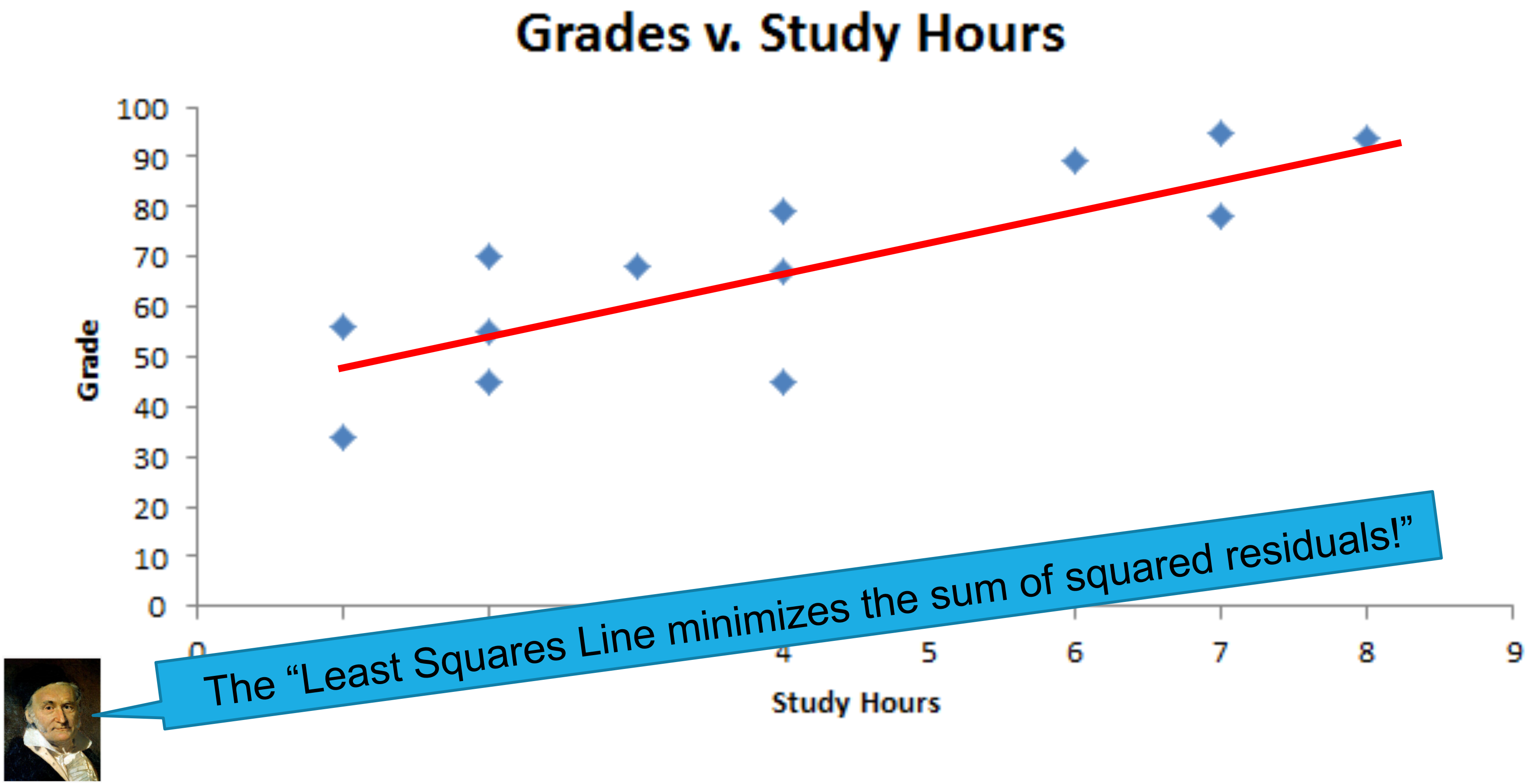
How Do We Estimate the Mean?



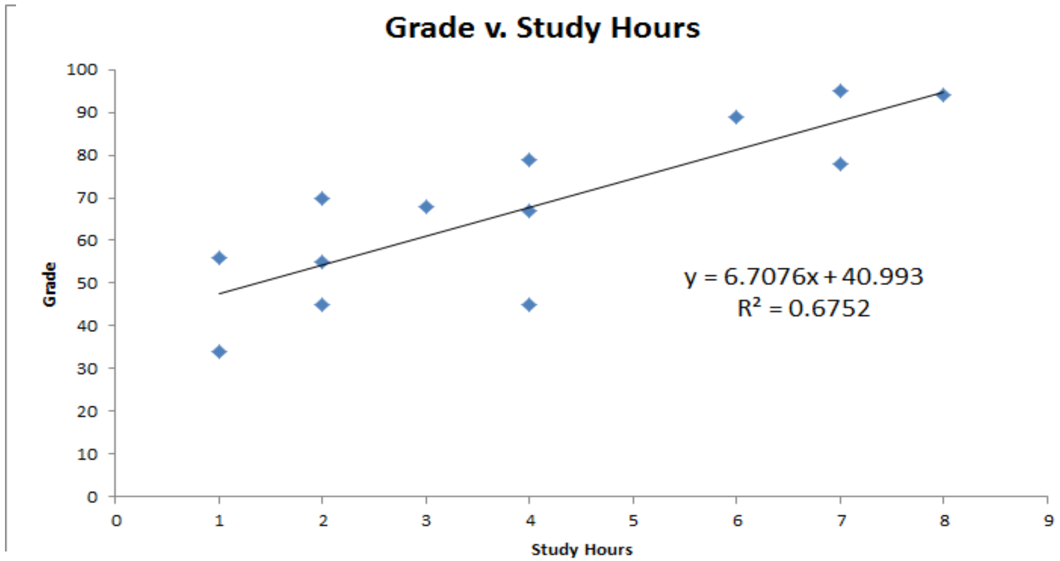
Observation: The sum of squared green residuals is greater than the sum of squared red residuals

Study Hours	1	1	2	2	2	3	4	4	4	6	7	7	8
Grades	34	56	45	70	55	68	67	79	45	89	95	78	94

How Do We Estimate the Mean? The Least Squares Principle



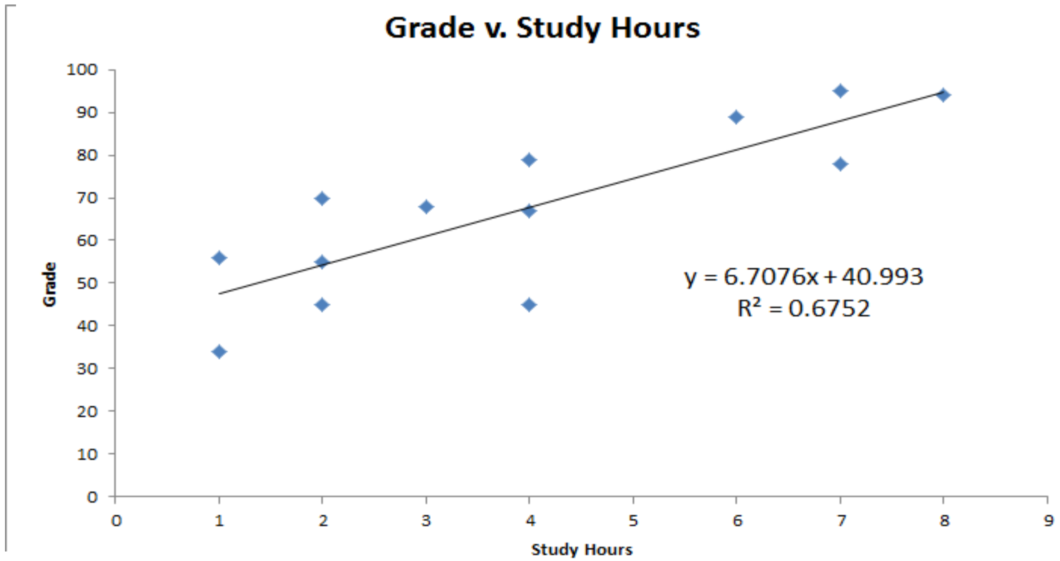
Example:
Grades vs. Study Hours



$$\mu\{Y|X\} = \beta_0 + \beta_1 X \rightarrow \hat{\mu}\{Y|X\} = \hat{\beta}_0 + \hat{\beta}_1 X = 40.993 + 6.7076\,X$$

The value of $\hat{\mu}\{Y|X_i\}$ is known a **FITTED OR PREDICTED VALUE**
The i^{th} **RESIDUAL** is defined to be $res_i = Y_i - \hat{\mu}\{Y|X_i\}$

Example: Grades vs. Study Hours

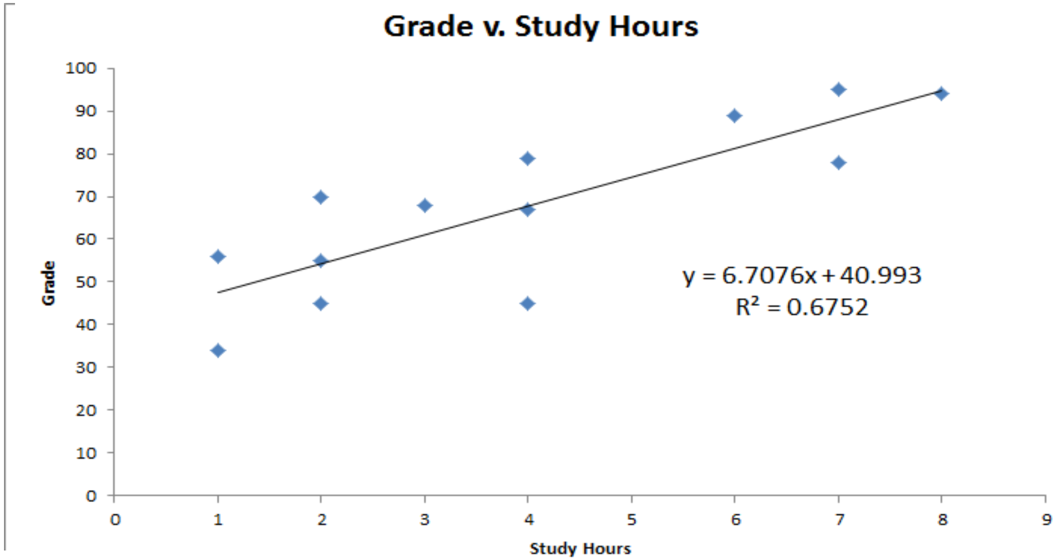


$$\mu\{Y|X\} = \beta_0 + \beta_1 X \ \rightarrow \ \hat{\mu}\{Y|X\} = \hat{\beta}_0 + \hat{\beta}_1 X = 40.993 + 6.7076\, X$$

What is the fitted value for X = 4 hours?
$\hat{\mu}\{Y|X = 4\} = 40.993 + 6.7076\,(4) = 67.8234$ points

"We estimate the mean score after studying for 4 hours to be 67.8234 points"

Example: Grades vs. Study Hours

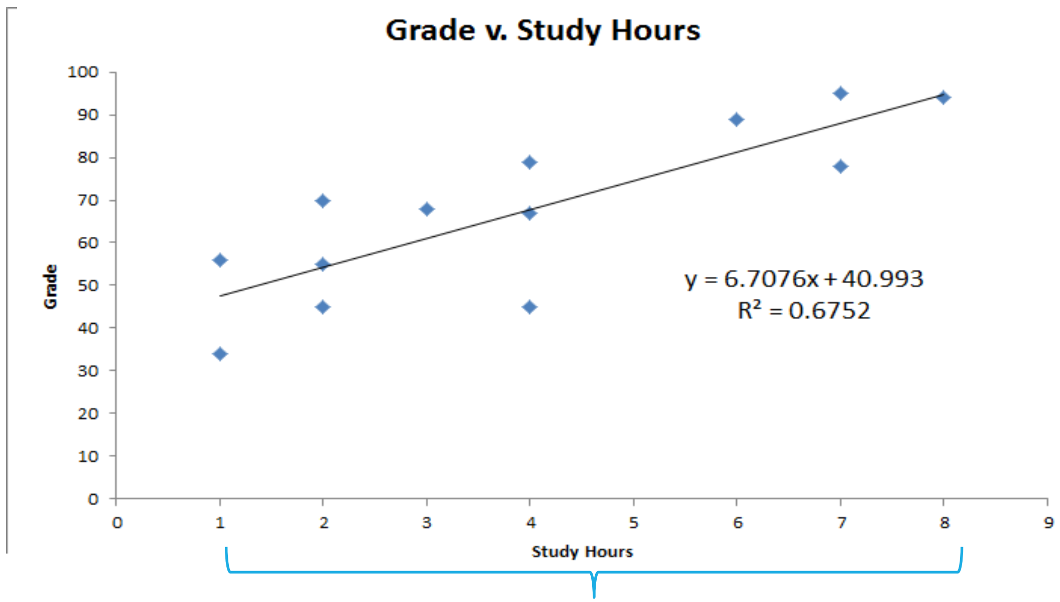


$$\mu\{Y|X\} = \beta_0 + \beta_1 X \rightarrow \hat{\mu}\{Y|X\} = \hat{\beta}_0 + \hat{\beta}_1 X = 40.993 + 6.7076\,X$$

What is the fitted value for X = 7 hours?
$\hat{\mu}\{Y|X = 7\} = 40.993 + 6.7076\,(6) = 87.946$ points

"We estimate the mean score after studying for 7 hours to be 87.946 points"

Extrapolation

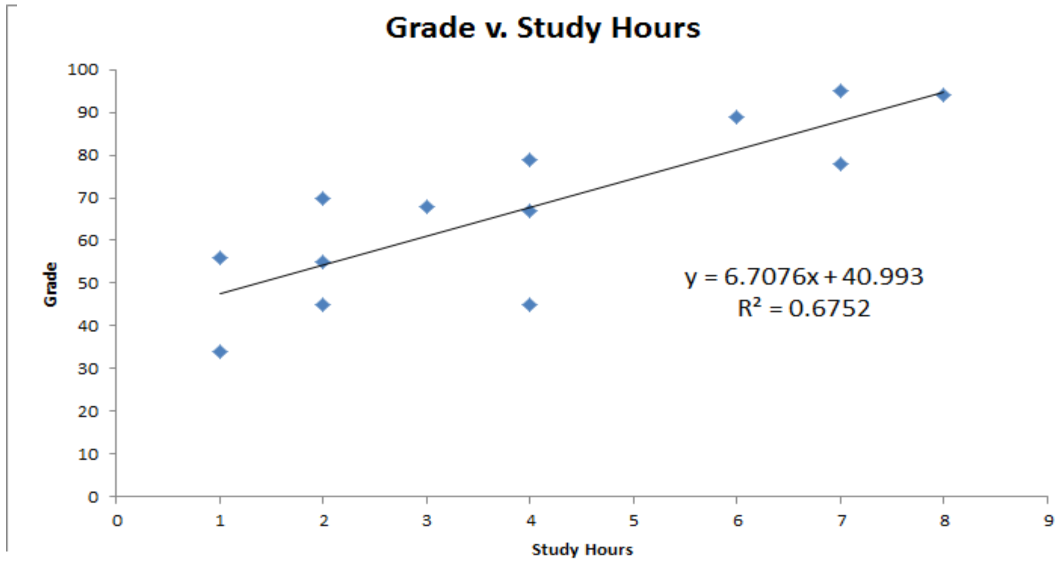


Predictions are only valid for values of X in the range of $X_1, X_1, \dots, X_n$

This least squares fit is only valid for study hours between 1 and 8 hours

Any prediction outside this range is known as an **EXTRAPOLATION**

Extrapolation

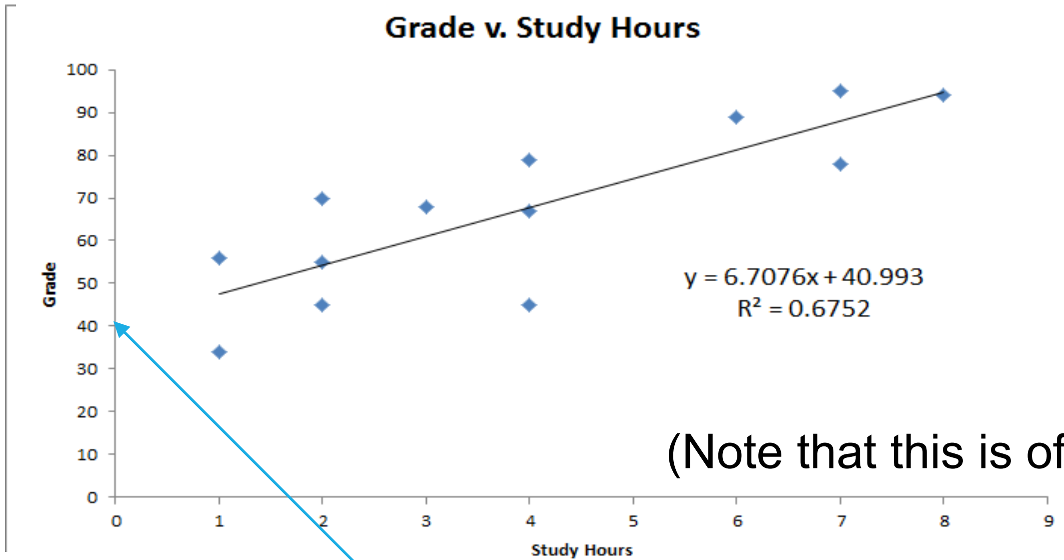


$$\mu\{Y|X\} = \beta_0 + \beta_1 X \ \rightarrow \ \hat{\mu}\{Y|X\} = \hat{\beta}_0 + \hat{\beta}_1 X = 40.993 + 6.7076\, X$$

What is the fitted value for X = 18 hours?

$\hat{\mu}\{Y|X = 4\} = 40.993 + 6.7076\,(18) = 161.4$ points

Interpretation



(Note that this is often an extrapolation)

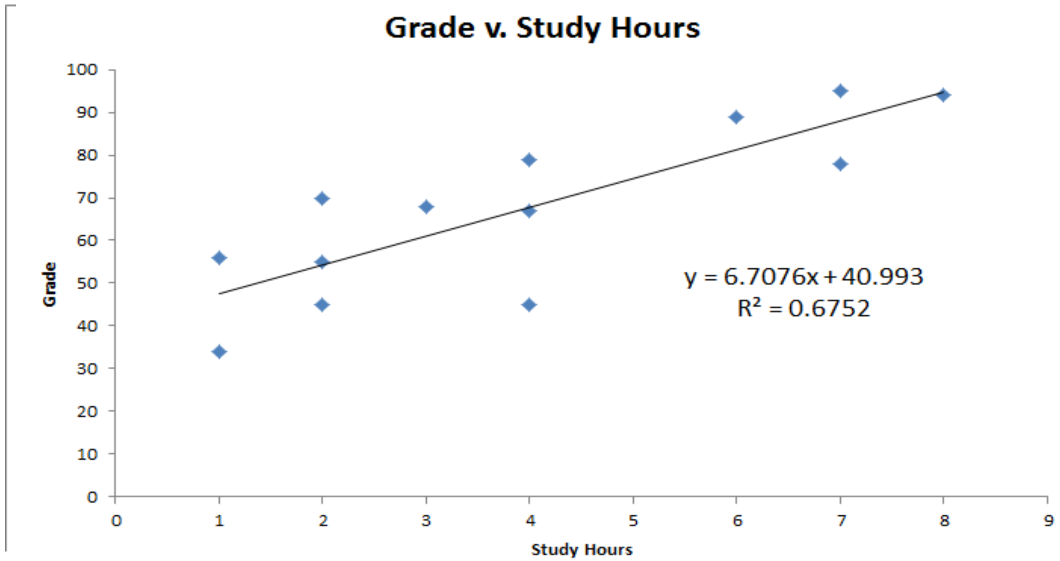
Interpreting $\hat{\beta}_0$ and $\hat{\beta}_1$ is an important part of simple linear regression

As $\hat{\mu}\{Y|X=0\} = \hat{\beta}_0 + \hat{\beta}_1(0) = \hat{\beta}_0$ $\rightarrow$ the intercept is predicted value at $X = 0$

Also, $\hat{\mu}\{Y|X=4\} - \hat{\mu}\{Y|X=3\} = \hat{\beta}_0 + \hat{\beta}_1(4) - \hat{\beta}_0 - \hat{\beta}_1(3) = \hat{\beta}_1$

$\rightarrow$ the slope is the change in the prediction for a 1 unit change in X

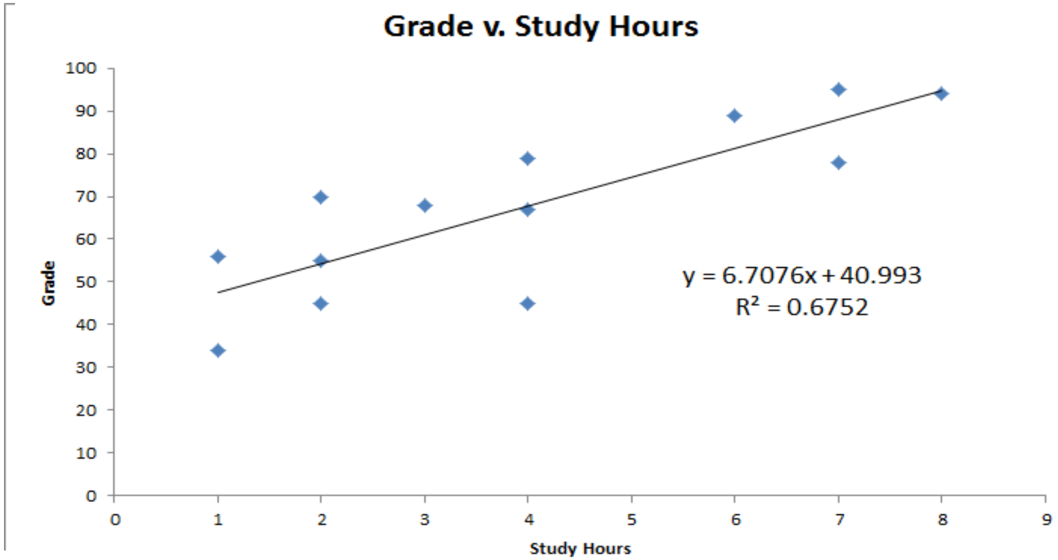
Interpretation



What is the predicted value for X = 0 hours?
$\hat{\mu}\{Y|X = 0\} = 40.993 + 6.7076\ (0) = 40.993$ points

We estimate that a 1 hour increase in study hours is associated with a 6.7076 points increase in predicted test grade (it is also ok to say “mean test grade”)

Bad Interpretations



We estimate that a 1 hour increase in study hours is associated with a 6.7076 points increase in test grade

A 1 hour increase in study hours leads to a 6.7076 points increase in test grade

Increasing study hours causes test grade to increase