CH. 13 – Linear Regression

PY 221 Statistics & Research Methods I

Dr. Valenti

Outline for Ch. 13 - Linear Regression

1. Review of prior tests we've learned
2. Overview of when to use linear regression
3. What is simple linear regression?
   - equation for the line of best fit
   - making predictions for an outcome (Y) from a given predictor value (X)
   - assessing whether individual predictors are significantly related to the outcome – return of the *t* statistic!
     - null and alternative hypotheses
4. Effect size (return of R^2!)
5. Multiple regression (time-permitting)

What do you remember?

- Use your handout to record the key differences and similarities between the statistical tests we've learned so far: t-test, one-way ANOVA, two-way ANOVA, and correlation.

REVIEW

- What do ***t*-tests** allow you to do?
  - examine whether the apparent differences between **two** groups on a quantitative outcome variable are real differences (vs. produced by chance)
  - examine whether the apparent changes in people's quantitative scores across **two** time points are real changes (vs. produced by chance)

Are there differences in tastiness ratings across these two fruits? Or are they equally tasty?

REVIEW

- What does **ANOVA** allow you to do?
  - examine whether apparent differences between **three or more** groups on a quantitative outcome variable are real differences (vs. produced by chance)

Are there differences in sweetness ratings across these four types of berries? Or are they equally sweet?

REVIEW

- What does **CORRELATION** allow you to do?
  - examine whether an apparent linear *relationship* between two quantitative variables is real (vs. produced by chance)

Does the average daily temperature in a given growing season relate to how sweet oranges are when they're ripe?

Comparison of the tests we've learned so far...

TEST	t-test	1-way ANOVA	2-way ANOVA	correlation
Type of *predictor*	1 qualitative/ categorical, with only 2 categories/levels (binary)			
Type of *outcome*	quantitative			
What does test look for?	Are there **differences in means** across groups?			
Null hyp.	$\mu_1 = \mu_2$ *OR* $\mu_1 - \mu_2 = 0$			

Comparison of the tests we've learned so far...

TEST	t-test	1-way ANOVA	2-way ANOVA	correlation
Type of *predictor*	1 qualitative/ categorical, with only 2 categories/levels (binary)	1 qualitative/ categorical, with 3+ categories/levels	2 qualitative/categorical, with 2+ levels	
Type of *outcome*	quantitative	quantitative	quantitative	
What does test look for?	Are there **differences in means** across groups?	Are there **differences in means** across groups?	Are there **differences in means** across groups of factor A? of factor B? And, do the differences across groups of A differ depending on B?	
Null hyp.	$\mu_1 = \mu_2$ *OR* $\mu_1 - \mu_2 = 0$	$\mu_1 = \mu_2 = \mu_3$ (assuming a 3-level predictor)	*Didn't cover in a formal way.*	

For correlation,
you do not need to specify which variable is *predictor* vs. *outcome*.

Comparison of the tests we've learned so far...

TEST	t-test	1-way ANOVA	2-way ANOVA	correlation
Type of *predictor*	1 qualitative/ categorical, with only 2 categories/levels (binary)	1 qualitative/ categorical, with 3+ categories/levels	2 qualitative/categorical, with 2+ levels	quantitative
Type of *outcome*	quantitative	quantitative	quantitative	quantitative
What does test look for?	Are there **differences in means** across groups?	Are there **differences in means** across groups?	Are there **differences in means** across groups of factor A? of factor B? And, do the differences across groups of A differ depending on B?	Is there a **linear relationship** between two variables?
Null hyp.	$\mu_1 = \mu_2$ *OR* $\mu_1 - \mu_2 = 0$	$\mu_1 = \mu_2 = \mu_3$ (assuming a 3-level predictor)	*Didn't cover in a formal way.*	$\rho = 0$

Outline for Ch. 13 - Linear Regression

1. Review of prior tests we've learned
2. Overview of when to use linear regression
3. What is simple linear regression?
   - equation for the line of best fit
   - making predictions for an outcome (Y) from a given predictor value (X)
   - assessing whether individual predictors are significantly related to the outcome – return of the t statistic!
     - null and alternative hypotheses
4. Effect size (return of R^2!)
5. Multiple regression (time-permitting)

Regression: for assessing relationships & making specific predictions

- What does (LINEAR) **REGRESSION** allow you to do?
  - Examine whether an apparent linear relationship between scores of **two or more** *quantitative* variables is real (vs. produced by chance)…

    …**in a way that also allows us to make specific predictions about one variable from values of the other variable(s).**

How are correlation and linear regression similar vs. different?

Regression: for assessing relationships & making specific predictions

Simple linear regression involves *one predictor* (and one outcome).

- Your book focuses on this type of regression.

Multiple linear regression involves *two or more predictors* (and one outcome).

- We'll touch on this briefly this week.

Regression: for making specific predictions

Sue is unable to take the final exam due to a serious illness. If we wanted to predict what her score would have been on the final, what could we do?

Student		Final Exam Score (Y)
Leslie		48
Jennifer		43
Ruthann		41
Tim		29
Lindsay		43
Carl		27
Sue		?

Regression: for making specific predictions

Sue is unable to take the final exam due to a serious illness. If we wanted to predict what her score would have been on the final, what could we do?

$r = .75$

	PREDICTOR	OUTCOME
Student	**Midterm Score (X)**	**Final Exam Score (Y)**
Leslie	43	48
Jennifer	50	43
Ruthann	48	41
Tim	37	29
Lindsay	49	43
Carl	39	27
Sue	48	**?**

$M = 44.8$ (out of 50)

$M = 38.5$ (out of 50)

Research questions that would involve regression - examples

1. Does the # of chronic health issues (e.g., diabetes) a person has predict the longevity of their flu symptoms?

2. Do PY 101 student evaluation scores predict the # of psychology courses students will take over their remaining years at BSC?

3. Does the number of new emails a professor receives in a day predict how many scoops of ice cream they eat that night?

Outline for Ch. 13 - Linear Regression

1. Review of prior tests we've learned
2. Overview of when to use linear regression
3. **What is simple linear regression?**
   - **equation for the line of best fit**
   - making predictions for an outcome (Y) from a given predictor value (X)
   - assessing whether individual predictors are significantly related to the outcome – return of the t statistic!
     - null and alternative hypotheses
4. Effect size (return of R^2!)
5. Multiple regression (time-permitting)

What is simple linear regression?

OUTCOME,Y (final exam)

PREDICTOR, X (midterm score)

- A way of predicting the value of one variable from another.
  - the simple regression model is *linear*
    - therefore, we describe the relationship between two variables using the equation of a straight line.
- Sample data for two variables (X and Y) are used (by JAMOVI) to calculate this equation, and the equation can then be used to make predictions about future data.

Regression equation for simple regression

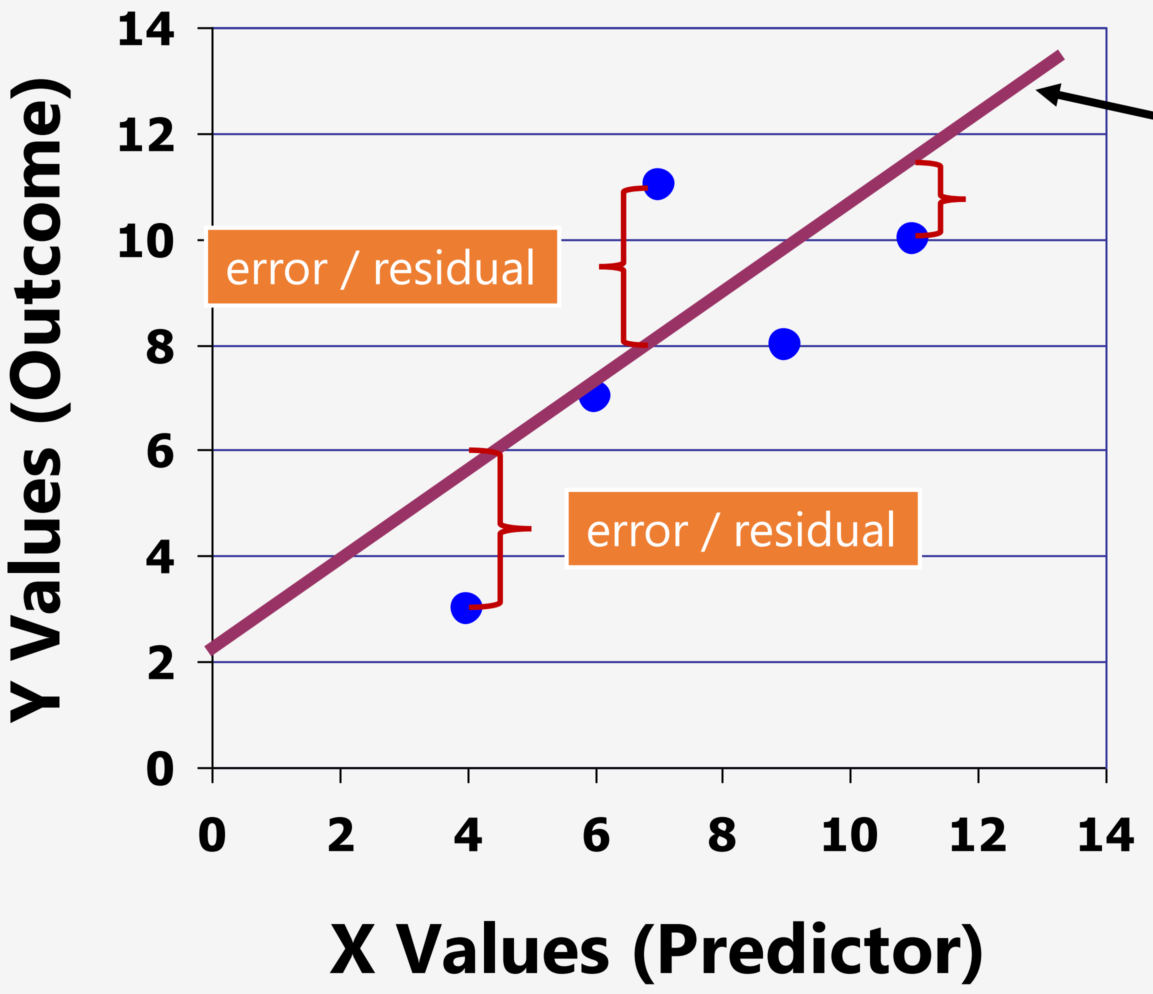
$\hat{Y}$ is the predicted value for the outcome variable Y

$$\hat{Y} = b_0 + b_1X_i$$

Equation of
a straight line
(*aka* the regression line,
aka the best fit line)

- b_1 (textbook uses just "b")
  - "**Regression coefficient**" for the predictor
  - Direction & strength of relationship btwn X & Y
  - **Slope** of the regression line (+, –, or 0)
    - Unlike r, b_1 is **not** restricted to falling btwn -1 and +1
- b_0 (textbook uses "a")
  - **Intercept** (value of Y when X = 0)
  - Point at which regression line crosses the Y-axis
- X_i
  - A value of X (the predictor) for which you want to predict the corresponding Y_i (the outcome)

The Line of Best Fit – example #1



This "best fit" line is represented with this regression equation:

$$\hat{Y} = b_0 + b_1 X$$

Y-intercept of line

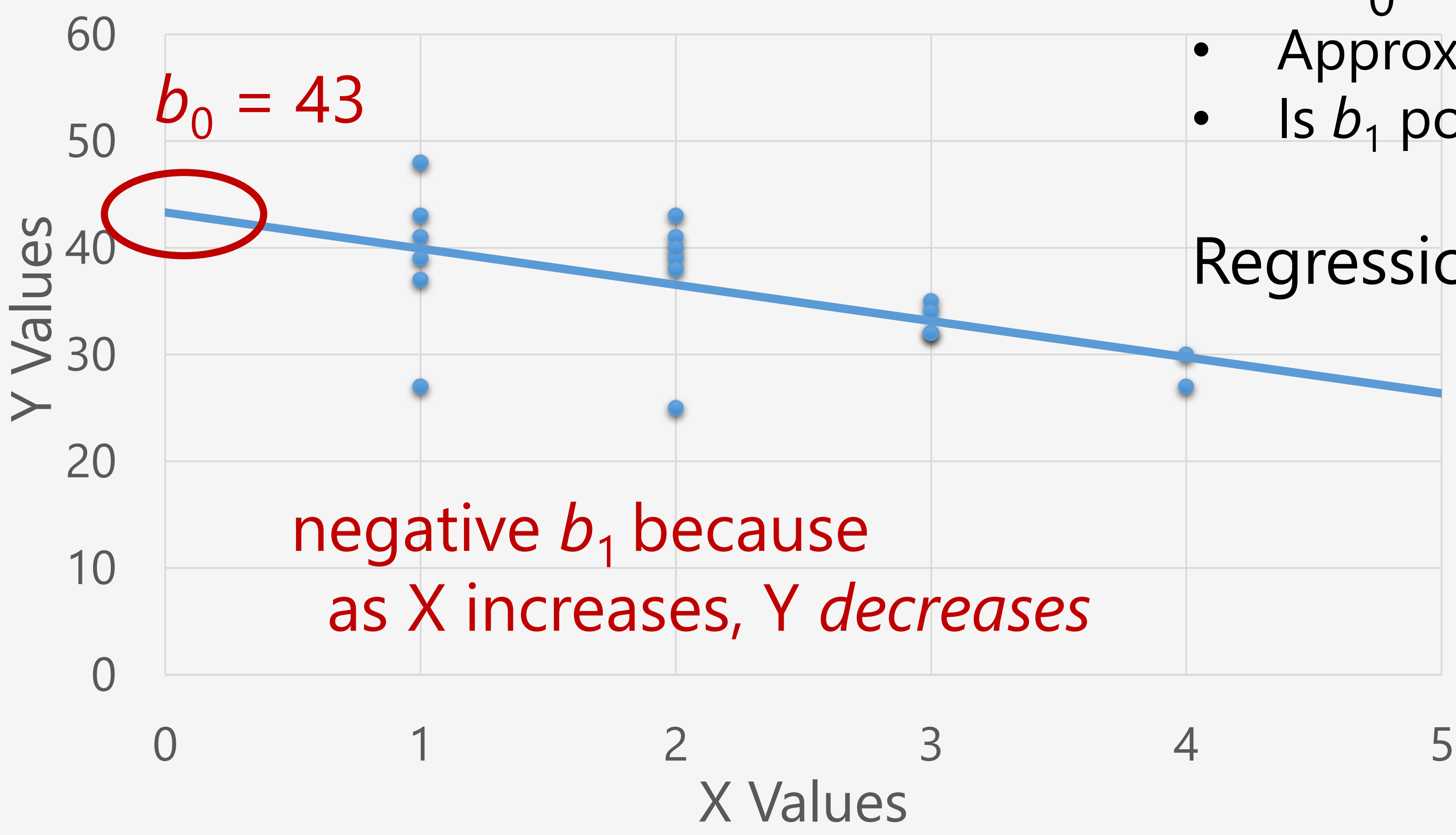
Slope of line

- Approximately what value will b_0 be?
- Is b_1 positive or negative?

$$\hat{Y} = 2 + 0.85X$$

FYI - We are not going to cover how to calculate the exact value of the slope.

The Line of Best Fit – example #2



$\hat{Y} = b_0 + b_1X$

- Approximately what value will b_0 be?
- Is b_1 positive or negative?

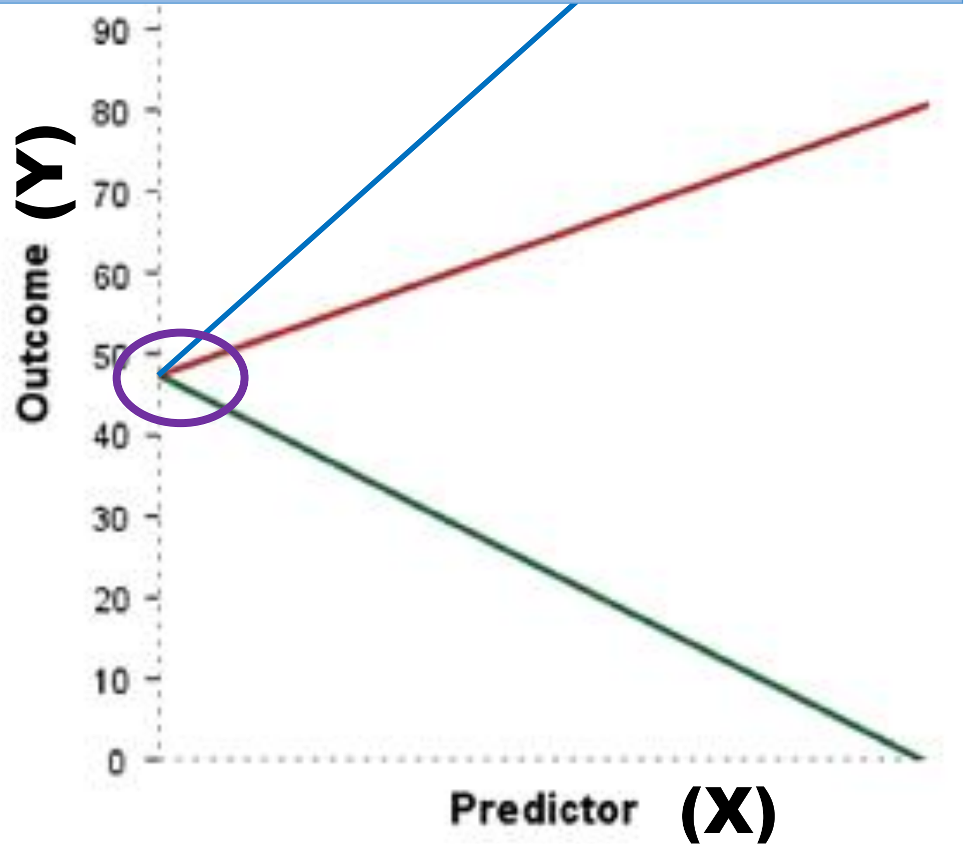
Regression Equation:

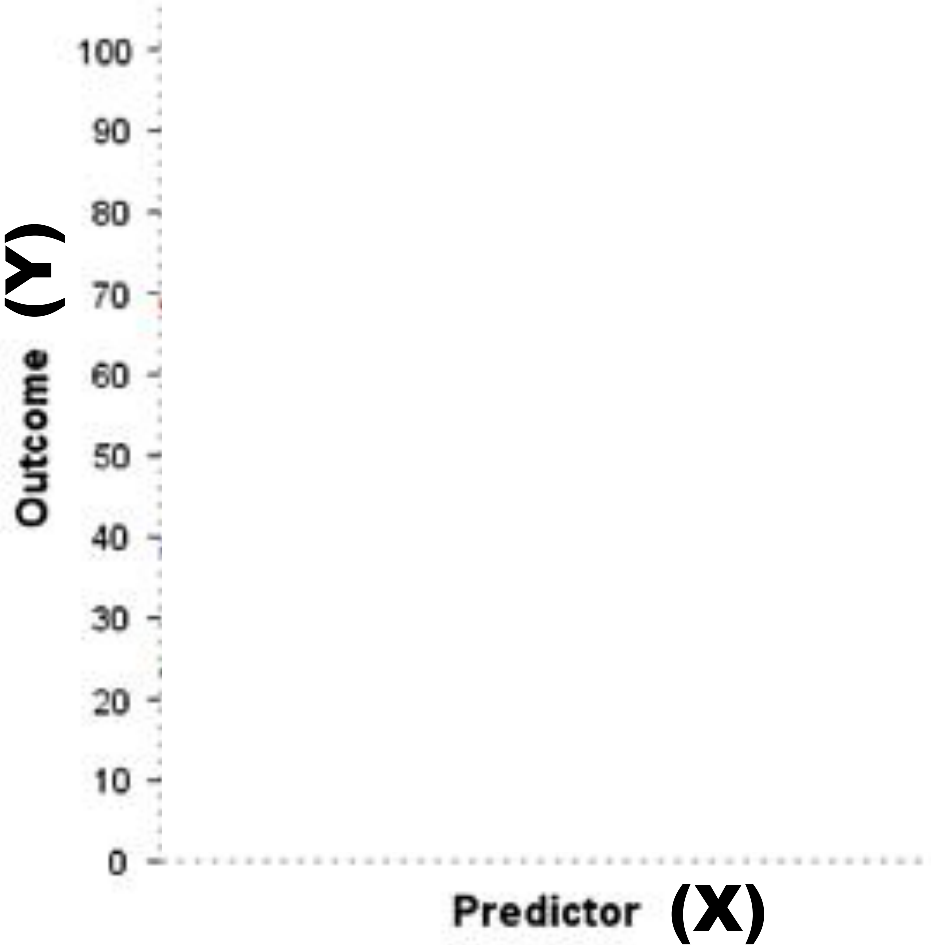
$\hat{Y} = 43 + -3.40X$

$\hat{Y} = 43 - 3.40X$

Intercepts (b_0) and Slopes (b_1)

All three regression lines have the
- same *intercepts* (b_0), but
- different *slopes* (b_1)

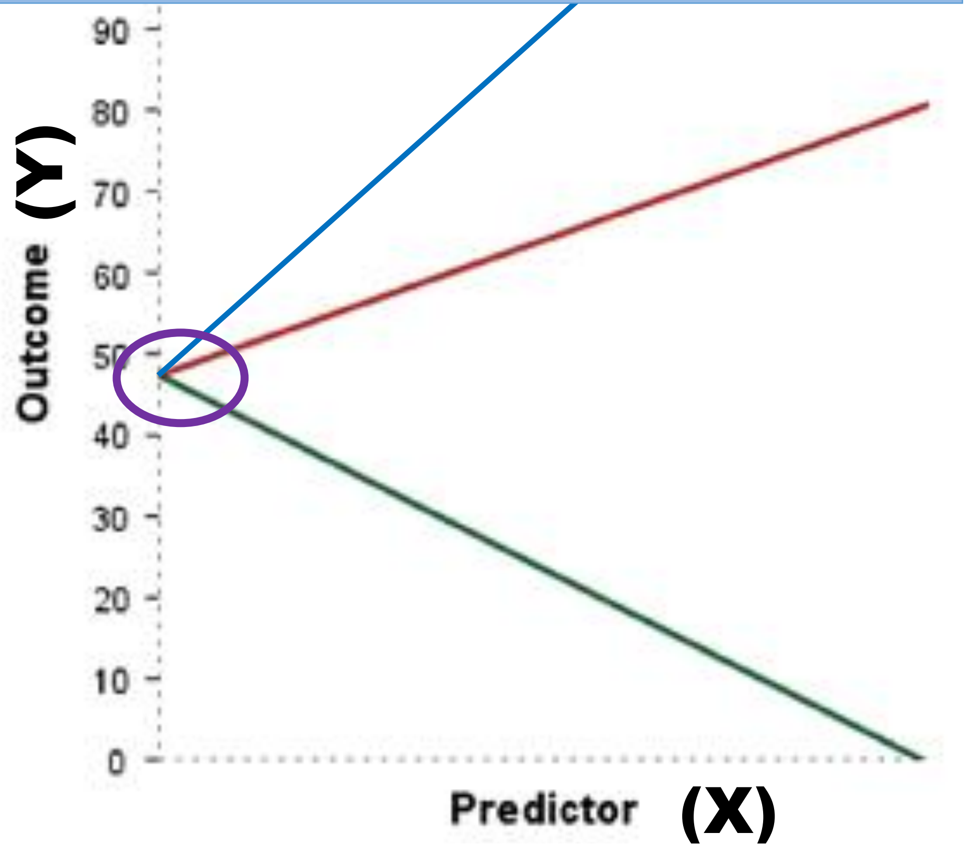


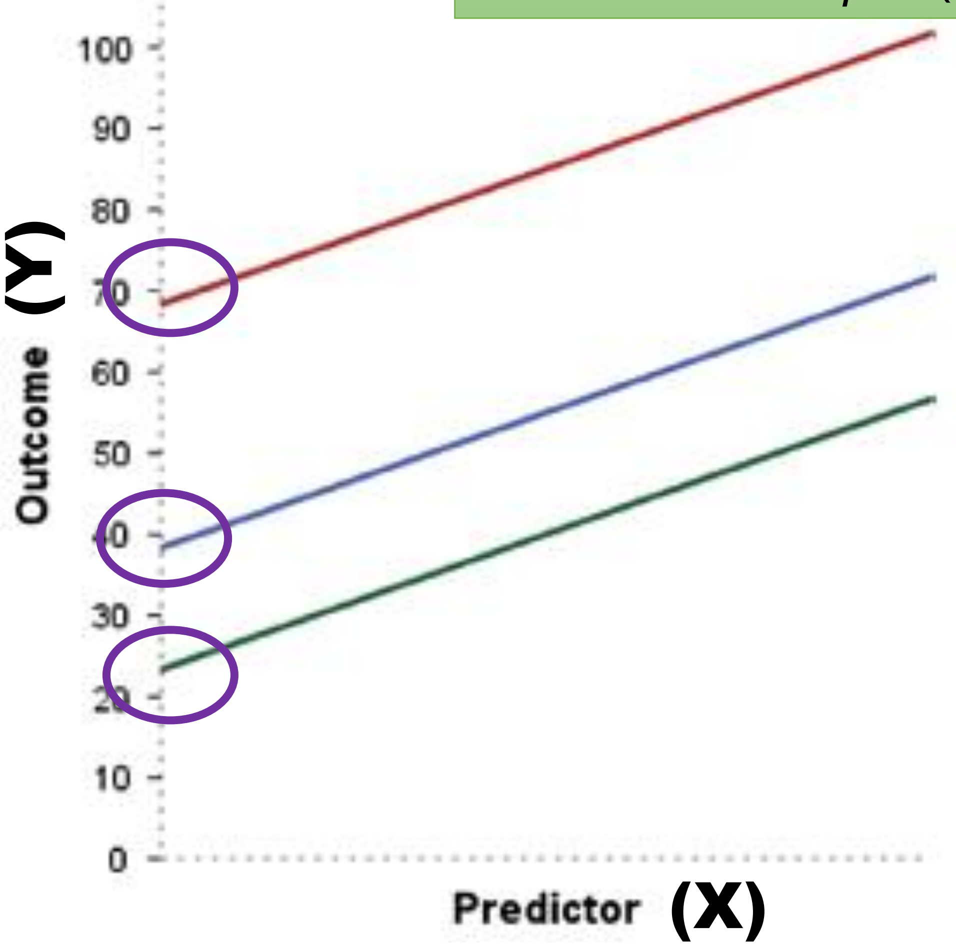


Intercepts (b_0) and Slopes (b_1)

All three regression lines have the
- same *intercepts* (b_0), but
- different *slopes* (b_1)

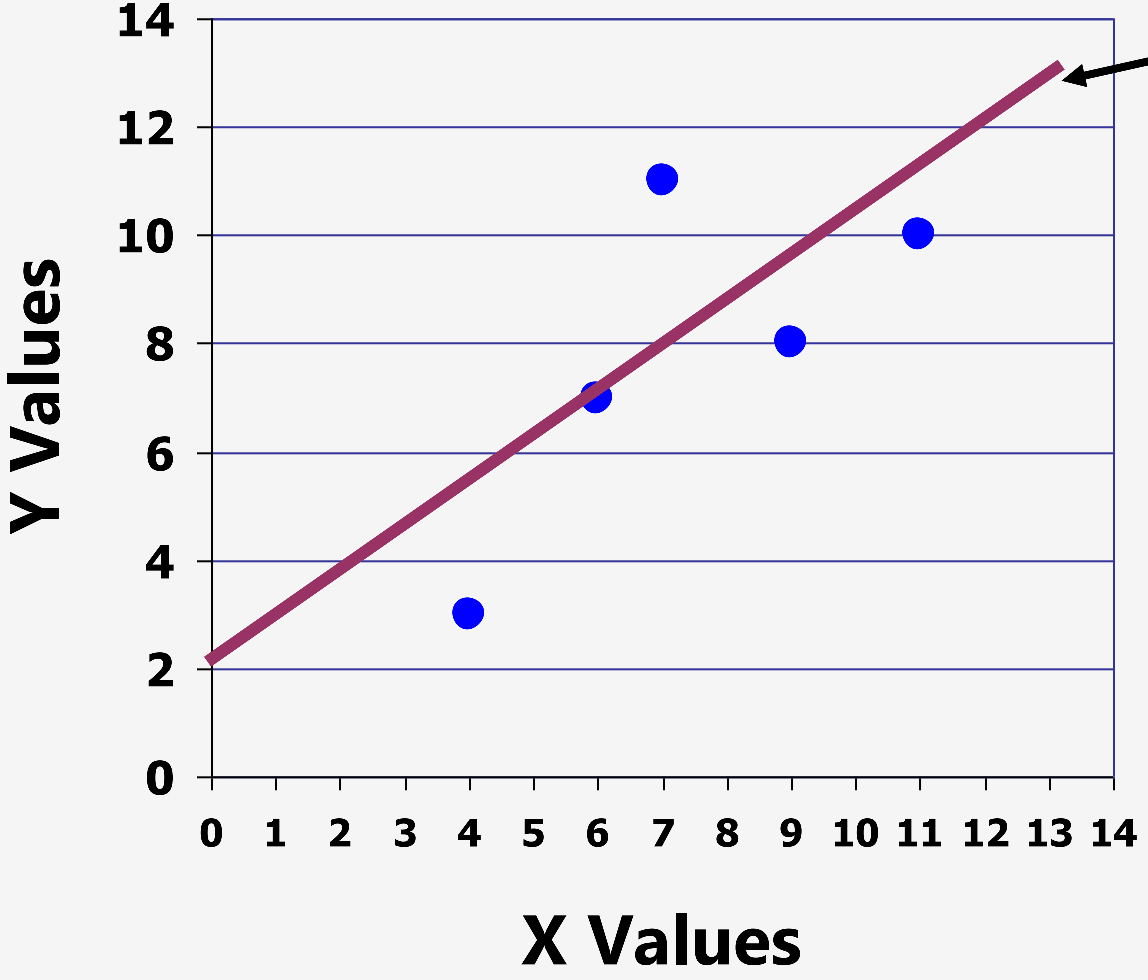
All three regression lines have
- different *intercepts* (b_0), but the
- same *slopes* (b_1)





We know the slopes are the same because the lines are *parallel.*

How to *interpret* the *value* of the slope (b_1) – ex #1



Regression equation for this "best fit" line: $\hat{Y} = 2 + \mathbf{0.85}X$

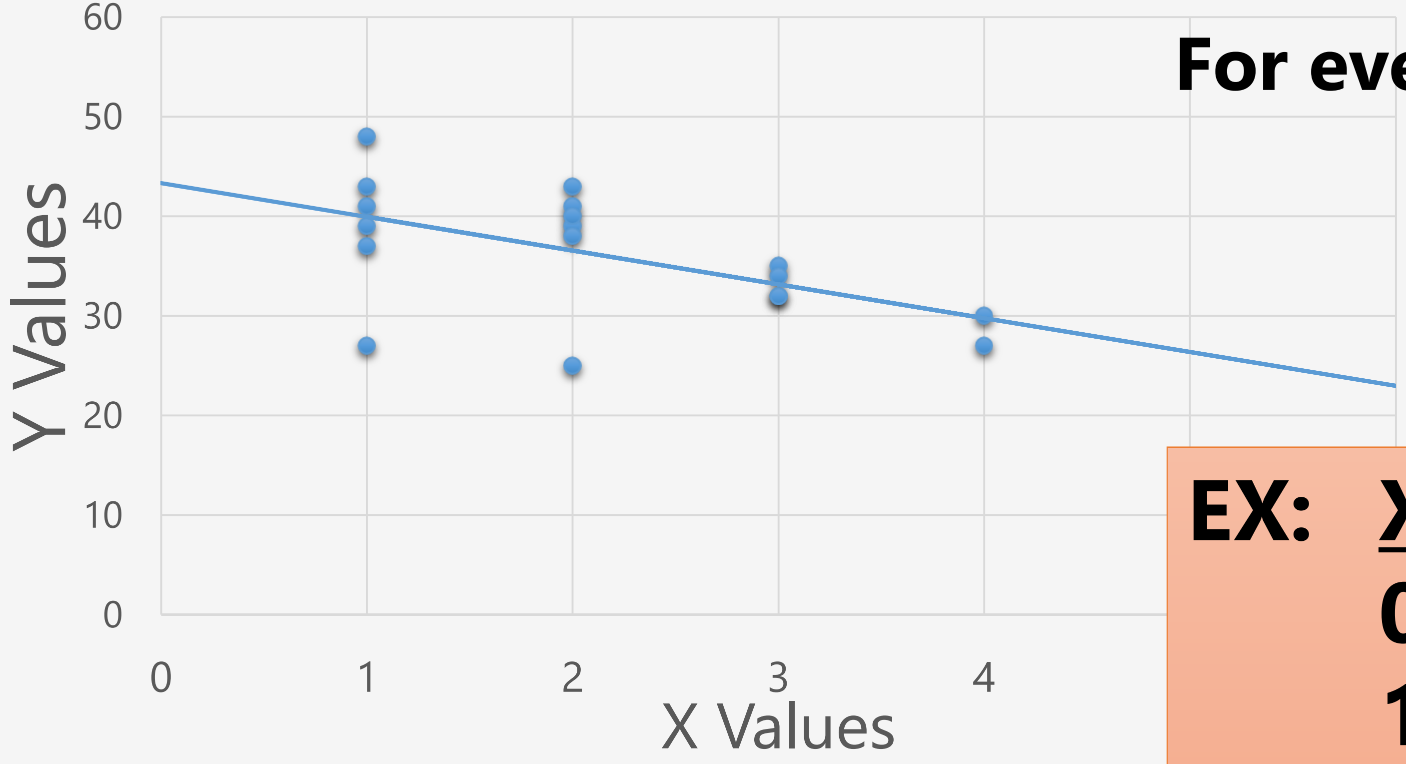
Q: We know the *positive* 0.85 indicates that X and Y are *positively* related, but what else does the 0.85 tell us?

For every 1 unit increase in X, Ŷ increases by 0.85 units.

EX:	X	Ŷ	
	0	*2.00*	An increase of .85 units
	1	*2.85*	An increase of .85 units
	2	*3.70*	An increase of .85 units
	3	*4.55*	
	etc…		

How to interpret slope (b_1) – ex #2

Regression Equation: $\hat{Y} = 43 - 3.40X$



For every 1 unit increase in X…

$\hat{Y}$ *decreases* by 3.40 units.

EX:	**X**	**$\hat{Y}$**	
	0	***43.0***	
			A decrease of 3.4 units
	1	***39.6***	
			A decrease of 3.4 units
	2	***36.2***	
	etc…		

Practice

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

Assume that a study was run and the data analyzed, and the regression coefficient is -2.0 and the intercept is 10.

1) Rewrite the equation in the box above, replacing the b_0 and b_1

2) What is the estimate of the outcome when the value of the predictor = 0?

A. 0 B. -2 C. 2 D. 10

3) Write an interpretation (translation) for the value of the slope:

For every 1 unit increase in the value of the predictor…

1. $\hat{Y} = 10 - 2.0X$
2. D 10
3. …the value of the outcome decreases by 2.0 units