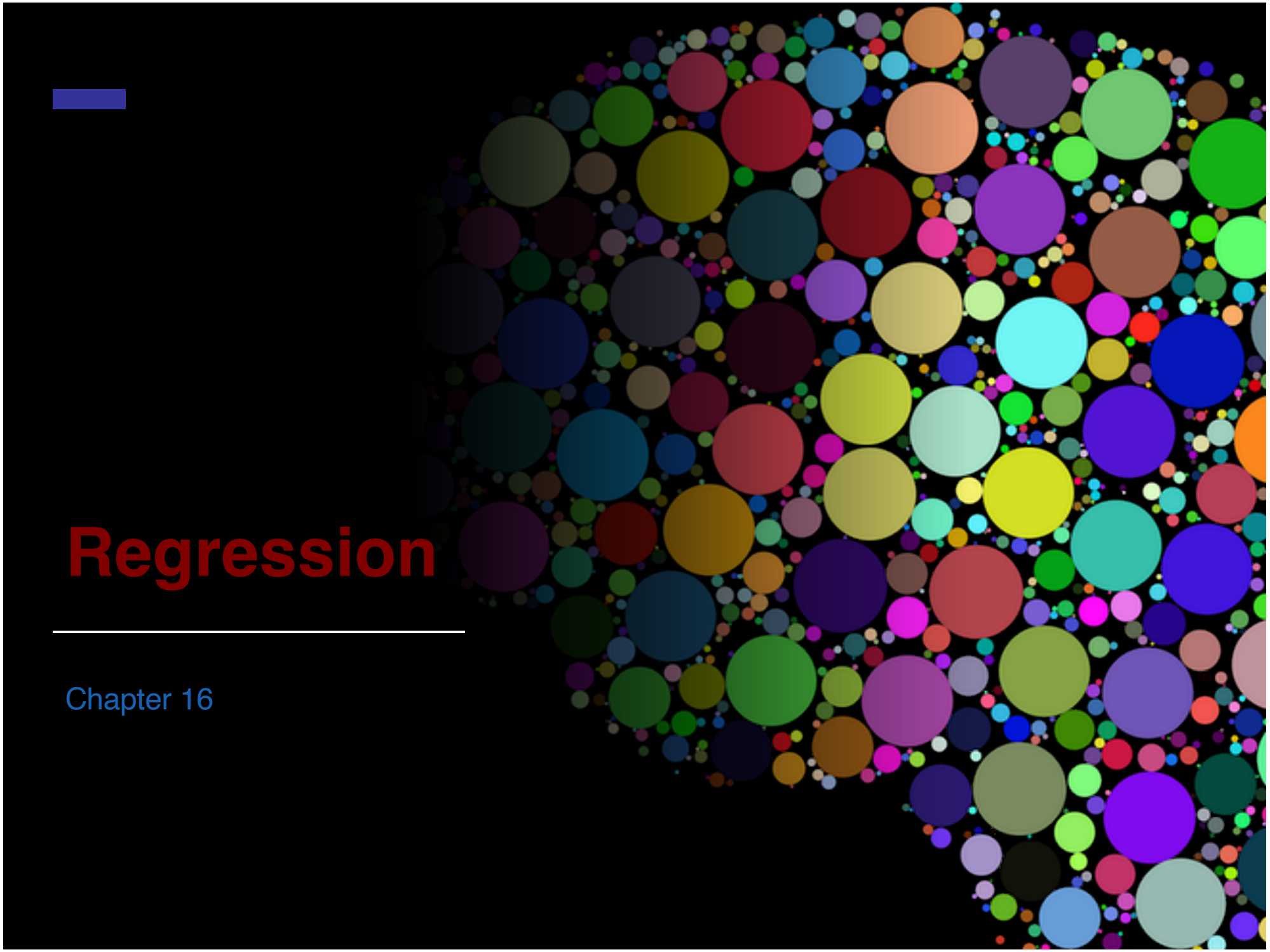


Regression

Chapter 16

Regression

> Builds on Correlation

> The difference is a question of prediction versus relation

- Regression predicts, correlation describes relations

Simple Linear Regression

> A statistical tool that lets us predict an individual's score on the DV based on the score on **one** IV
  - Multiple linear regression uses **more than one IV**

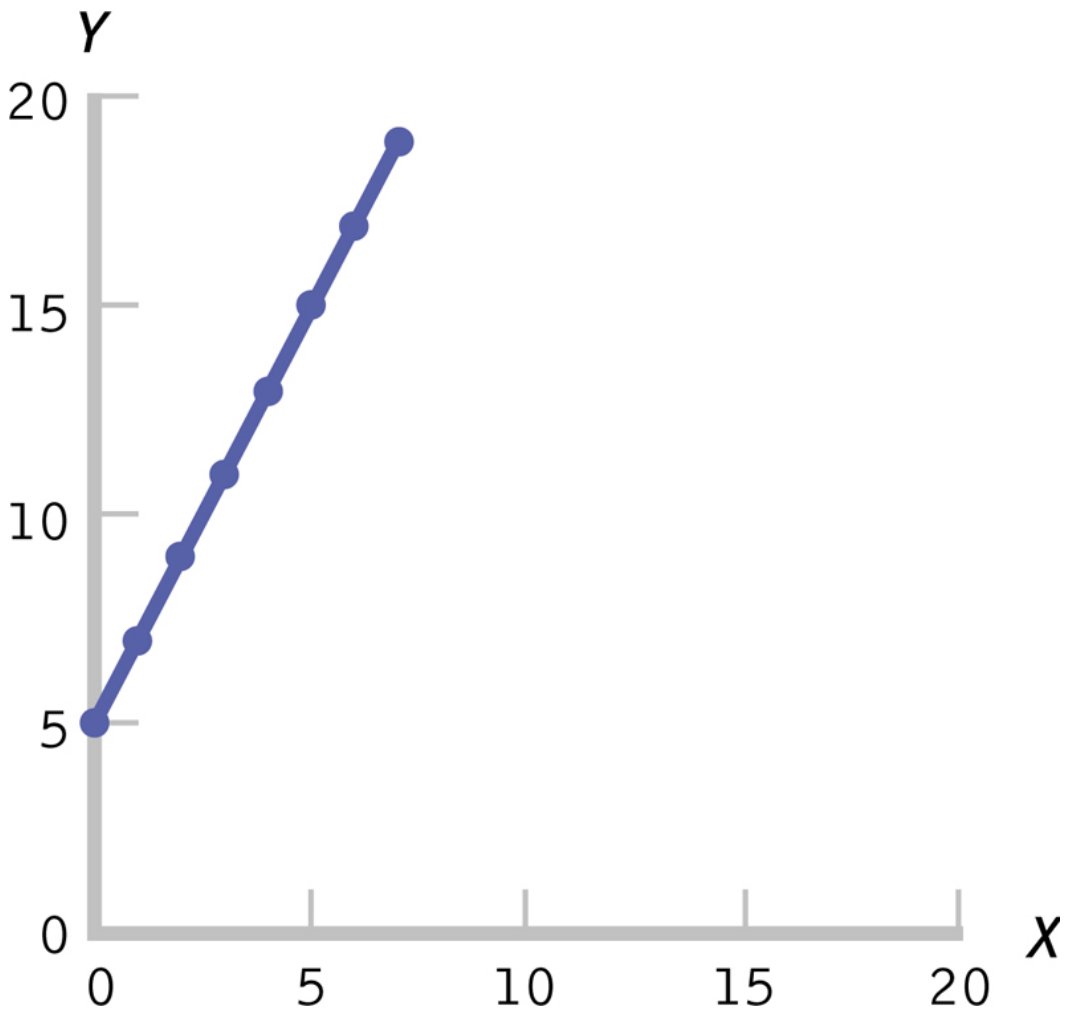
Linear Regression

> a = Intercept: predicted value of *Y* when *X* is equal to 0
  - Mean of Y

> b = Slope: the amount that *Y* is predicted to increase for an increase of 1 in *X*

$$\hat{Y} = a + bX$$

The Equation for a Line



Standardized regression coefficient

> Statisticians love to standardize things.

> *Beta* = standardized coefficient

- It's a z score of the slope value
- WHY? Comparisons.
- Beta = r with ONE predictor!

Assumptions

> Random selection

> X and Y are at least scale variables

> X and Y are both normal

> Homoscedasticity (for real this time!)

- Each variable must vary approximately the same at each point of the other variable.
- See scatterplot for UFOs, megaphones, snakes eating dinner.

Hypothesis Steps

> Step 1:

- Population: people where IV *does not predict* DV (i.e. b = 0)
- Sample: people where IV *does predict* DV

Hypothesis Steps

> Step 2:

> Null: No relationship between pronouns and social words (b = 0)

> Research: Relationship between pronouns and social words (b/=0)

> (just an example, using data from RP3 in class)

Hypothesis Steps

> Step 3:

> List regression equation

- Notice relationship between *beta* and *r*

> List the *df* = N – 2

Hypothesis Steps

> Get the equation:

- output = lm(*DV* ~ *IV*, data = data)
- summary(output)

Hypothesis Steps

> Get beta!

> Install the QuantPsyc library.

> library(QuantPsyc)

> lm.beta(model)

Hypothesis Steps

> Step 4

- List the cut off score $p < .05$
- qt(.05/2, df, lower.tail = F)

Hypothesis Steps

> Step 5:

- List the *t* value
- (don’t use the constant)

Hypothesis Steps

> Step 6:

- Reject or fail to reject!

Hypothesis Steps

> Effect size

- R^2 = proportionate reduction in error
- Same rules as ANOVA applies
  - >S = .01
  - >M = .06
  - >L = .14

> It's at the bottom of the output.

- This effect is for ALL predictors.

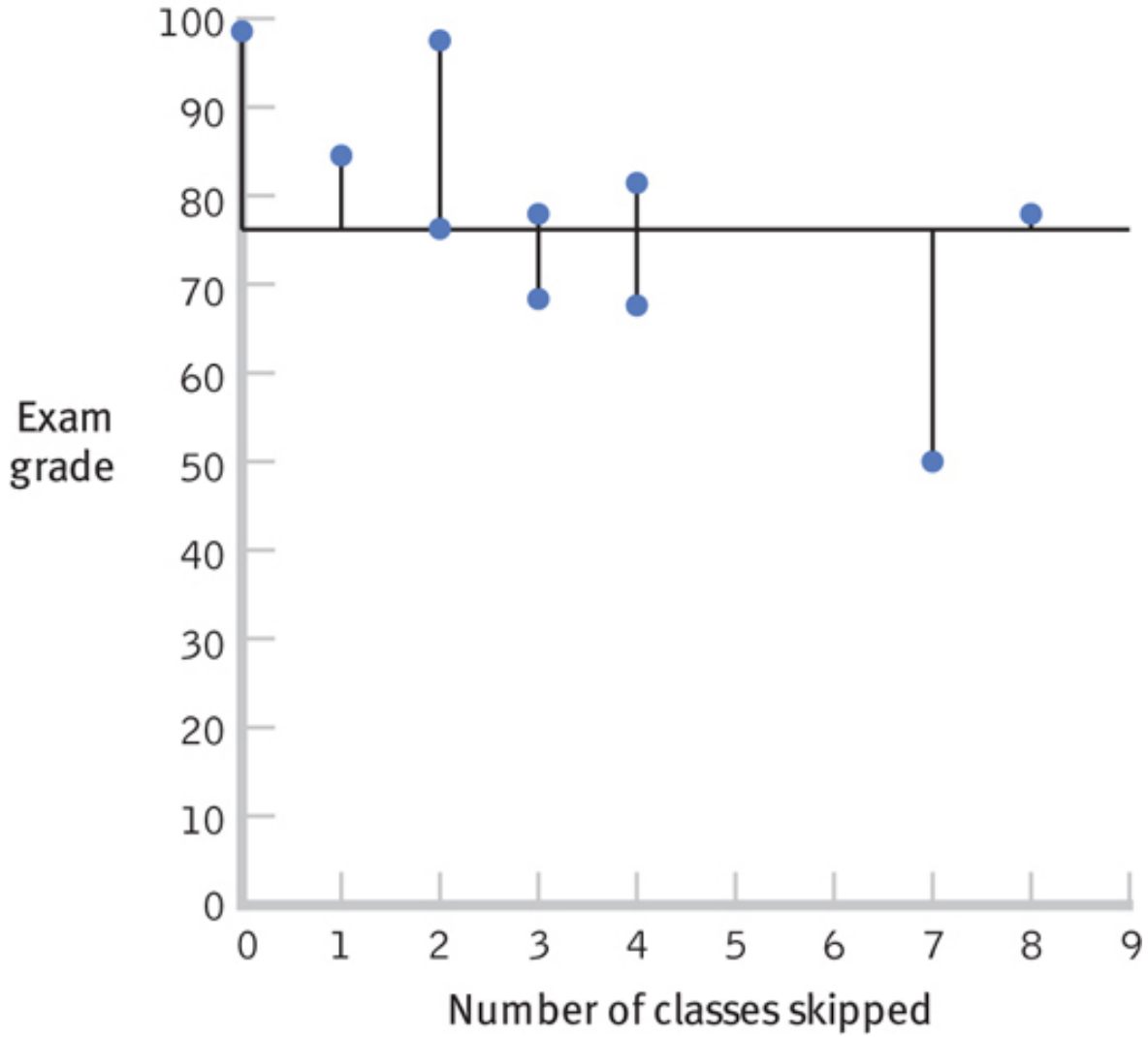
Proportionate Reduction in Error

> Also called the coefficient of determination

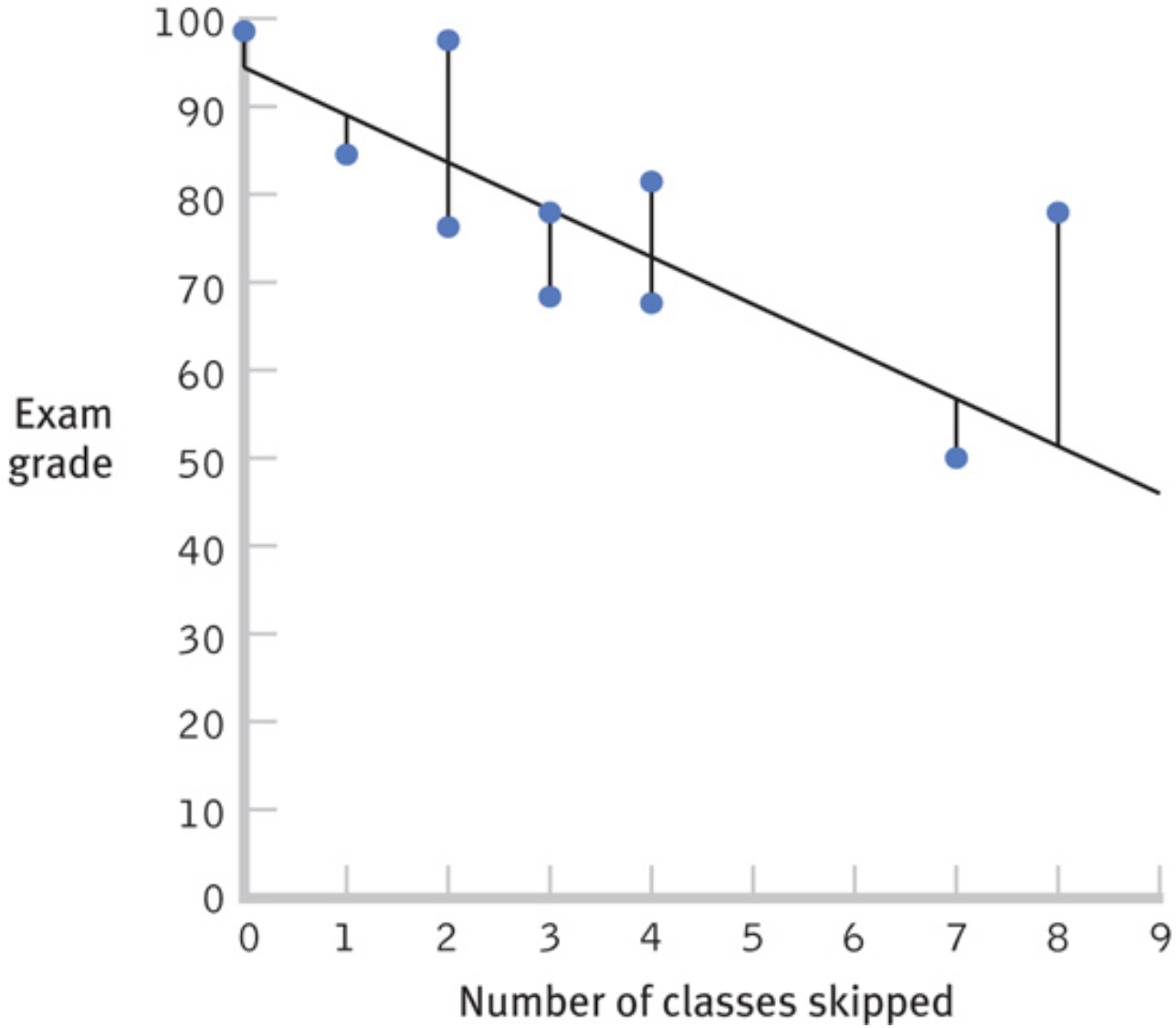
- Quantifies how much more accurate our predictions are when we use the regression line instead of the mean as a prediction tool

> Explaining what R^2 actually means

Visualizing Error



Visualizing Error



Regression Issues

> Standard error of the estimate

> Regression to the mean

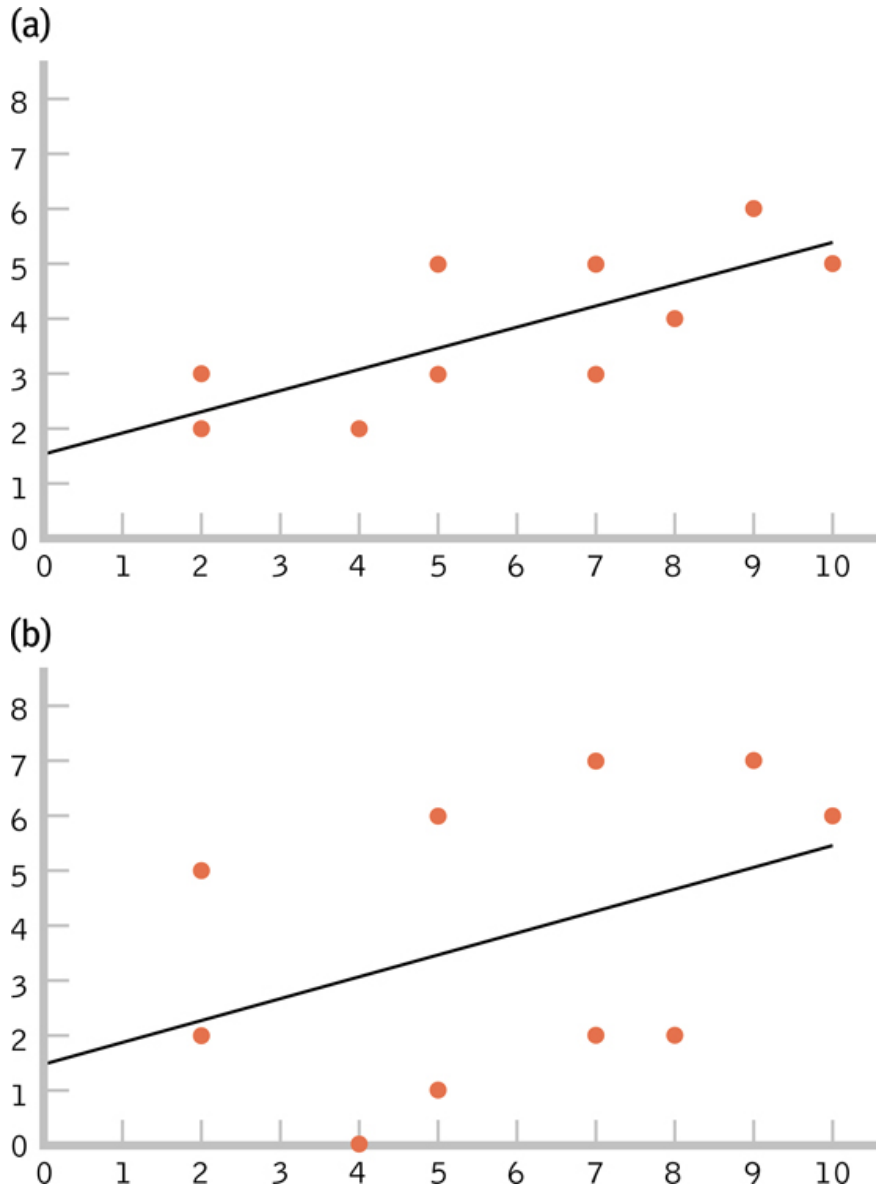
Interpretation and Prediction

> Standard error of the estimate

- Indicates the typical distance between the regression line and the actual data points
- Often called least squared error

The Standard Error of the Estimate

Which figure makes the stronger prediction? Why?



> Regression to the mean

- The patterns of extreme scores

Regression to the Mean

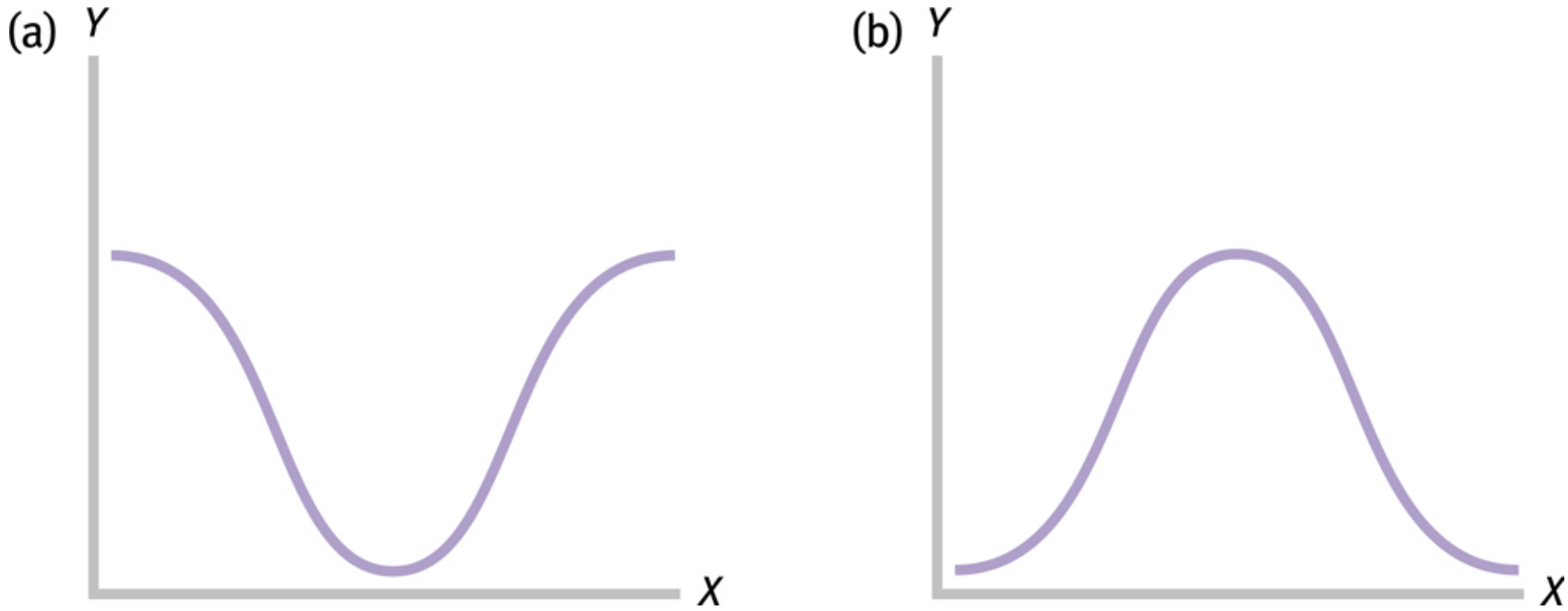


TABLE 16-5. Regression to the Mean: Investing

Bernstein (1996) presented these data from *Morningstar,* an investment publication, demonstrating regression to the mean in action. Notice that the category that showed the highest performances during the first time period (e.g., international stocks) had declined by the second time period, whereas the category with the poorest performances in the first time period (e.g., aggressive growth) had improved by the second time period.

5 Years to Objective	5 Years to March 1989	March 1994
International stocks	20.6%	9.4%
Income	14.3%	11.2%
Growth and income	14.2%	11.9%
Growth	13.3%	13.9%
Small company	10.3%	15.9%
Aggressive growth	8.9%	16.1%
Average	13.6%	13.1%

Other Regression Techniques

> Multiple Regression

> Stepwise Regression

> SEM

Multiple Regression

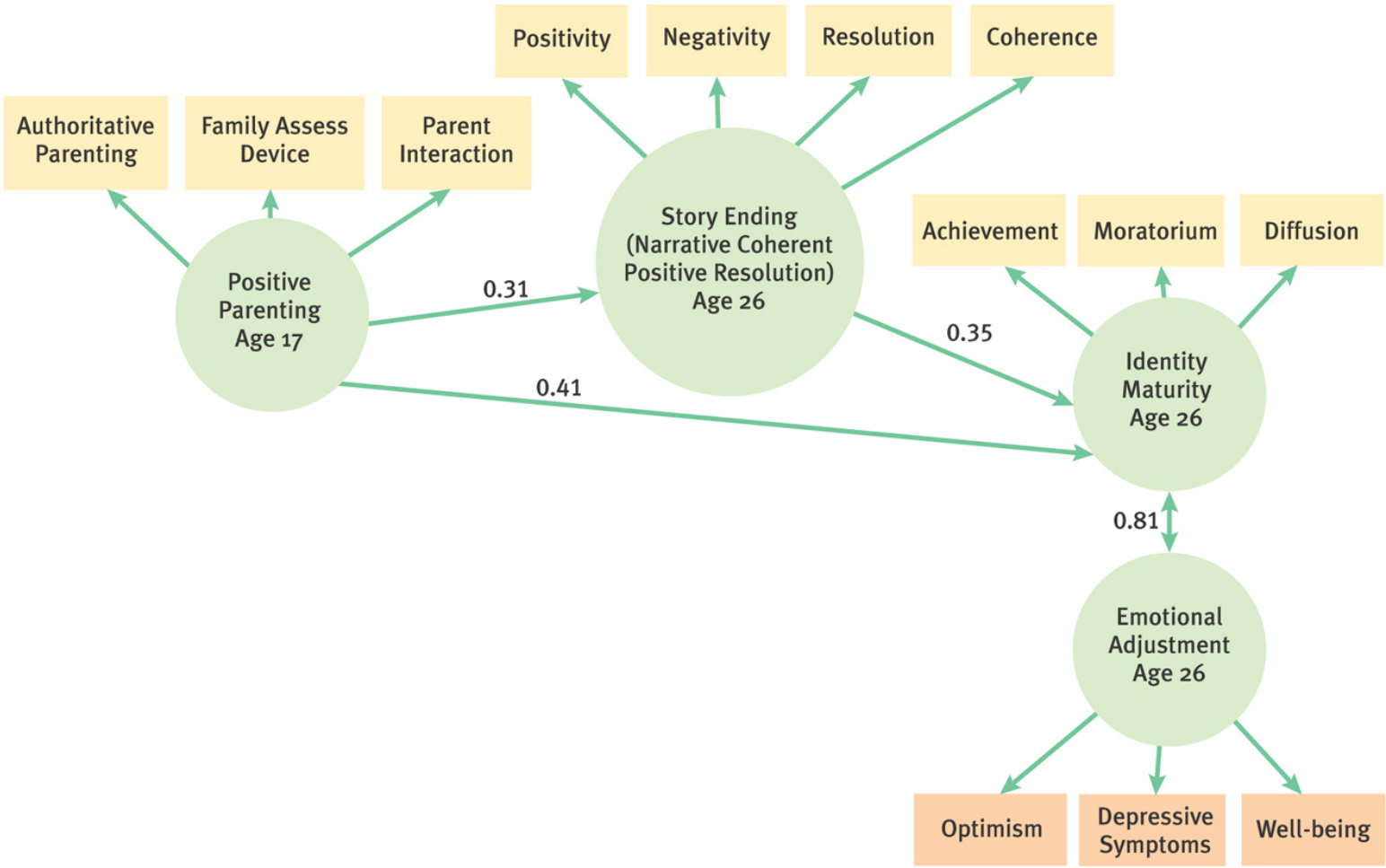
> A statistical technique that includes two or more predictor variables in a prediction equation

Stepwise Regression

- A type of multiple regression in which computer software determines the order in which IVs are included in the equation.
- It is the default in many computer software programs.
- The strength of using stepwise regression is its reliance on data, rather than theory— especially when a researcher is not certain of what to expect in a study.

Structural Equation Modeling (SEM)

- A statistical technique that quantifies how well sample data “fit” a theoretical model that hypothesizes a set of relations among multiple variables.
- SEM encourages researchers to think of variables as a series of connections.


Positivity
Negativity
Resolution
Coherence
Authoritative Parenting
Family Assess Device
Parent Interaction
Story Ending (Narrative Coherent Positive Resolution) Age 26
Achievement
Moratorium
Diffusion
Positive Parenting Age 17
0.31
0.35
0.41
Identity Maturity Age 26
0.81
Emotional Adjustment Age 26
Optimism
Depressive Symptoms
Well-being