Lecture 11: Chapter 5, Section 3 Relationships between Two Quantitative Variables; Correlation

- Display and Summarize
- Correlation for Direction and Strength
- Properties of Correlation
- Regression Line

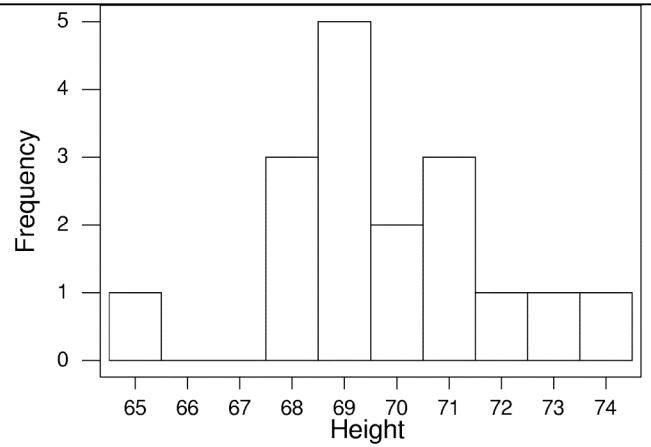
Looking Back: *Review*

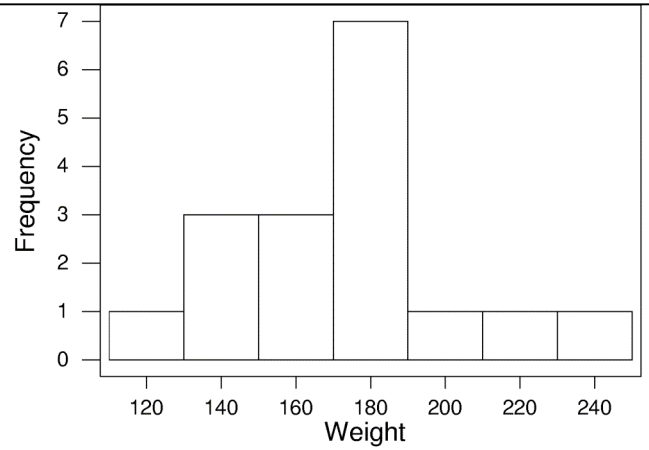
- **4 Stages of Statistics**
  - Data Production (discussed in Lectures 1-4)
  - Displaying and Summarizing
    - Single variables: 1 cat,1 quan (discussed Lectures 5-8)
    - Relationships between 2 variables:
      - Categorical and quantitative (discussed in Lecture 9)
      - Two categorical (discussed in Lecture 10)
      - Two quantitative
  - Probability
  - Statistical Inference

Example: *Two Single Quantitative Variables*

- **Background**: Data on male students' heights and weights:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
height	17	69.765	69.000	69.800	2.137	0.518
weight	17	170.59	175.00	169.33	28.87	7.00





- **Question:** What do these tell us about the relationship between male height and weight?
- **Response:**

Definition

- **Scatterplot** displays relationship between 2 quantitative variables:
  - Explanatory variable (x) on horizontal axis
  - Response variable (y) on vertical axis

Example: *Explanatory/Response Roles*

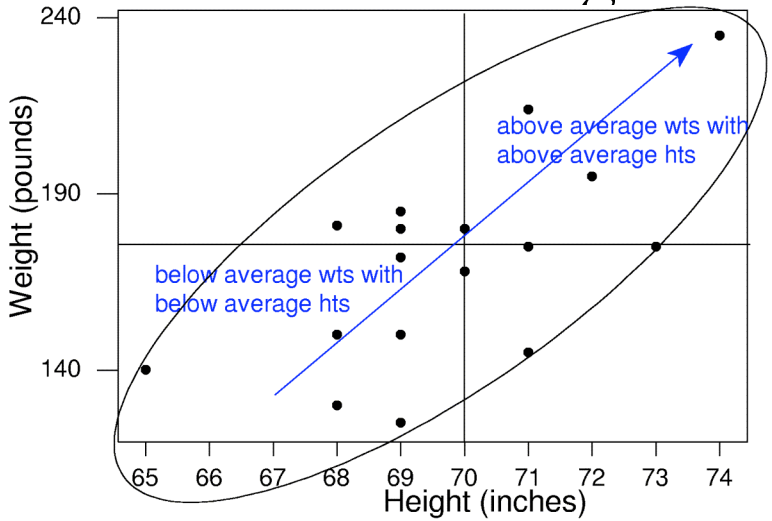
- **Background**: We're interested in the relationship between male students' heights and weights.
- **Question:** Which variable should be graphed along the horizontal axis of the scatterplot?
- **Response:**

Definitions

- **Form:** relationship is linear if scatterplot points cluster around some straight line
- **Direction:** relationship is
  - positive if points slope upward left to right
  - negative if points slope downward left to right

Example: *Form and Direction*

- **Background**: Scatterplot displays relationship between male students' heights and weights.



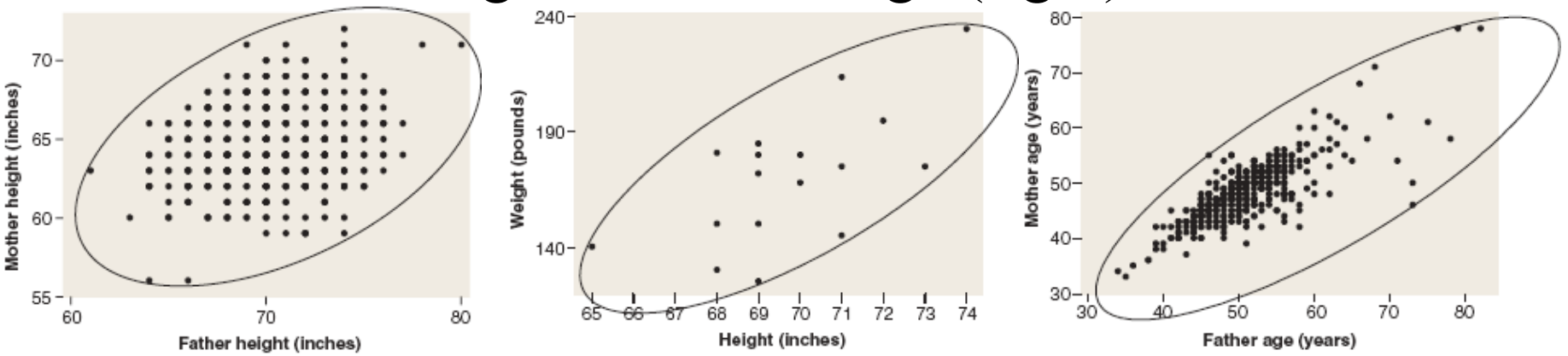
- **Question:** What are the form and direction of the relationship?
- **Response:** Form is _____ direction is _____

Strength of a Linear Relationship

- **Strong:** scatterplot points tightly clustered around a line
  - Explanatory value tells us a lot about response
- **Weak:** scatterplot points loosely scattered around a line
  - Explanatory value tells us little about response

Example: *Relative Strengths*

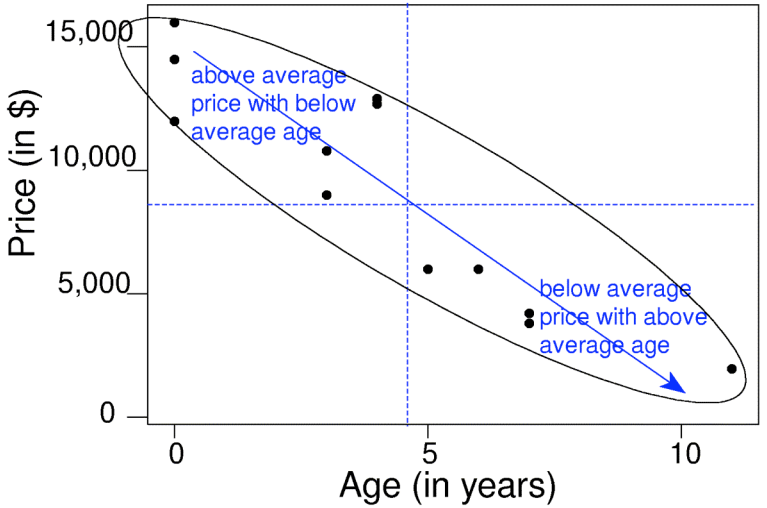
- **Background**: Scatterplots display:
  - mothers' ht. vs. fathers' ht. (left)
  - males' wt. vs. ht. (middle)
  - mothers' age vs. fathers' age (right):



- **Question:** How do relationships' strengths compare? (Which is strongest, which is weakest?)
- **Response:** Strongest is on weakest is on

Example: *Negative Relationship*

- **Background**: Scatterplot displays price vs. age for 14 used Pontiac Grand Am's.



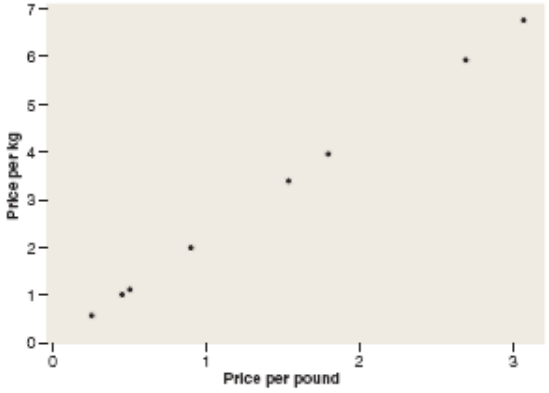
- **Questions:**
  - Why should we expect the relationship to be negative?
  - Does it appear linear? Is it weak or strong?
- **Responses:**
  -
  -

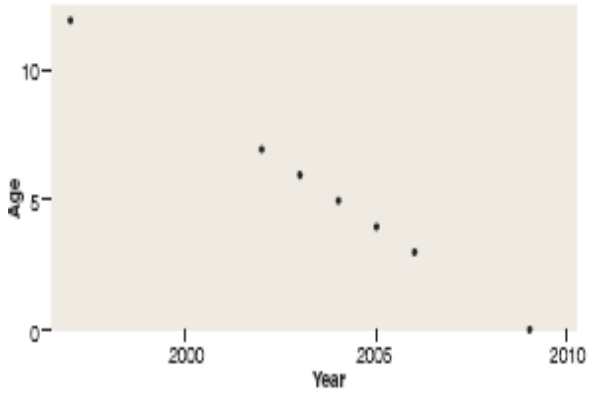
Definition

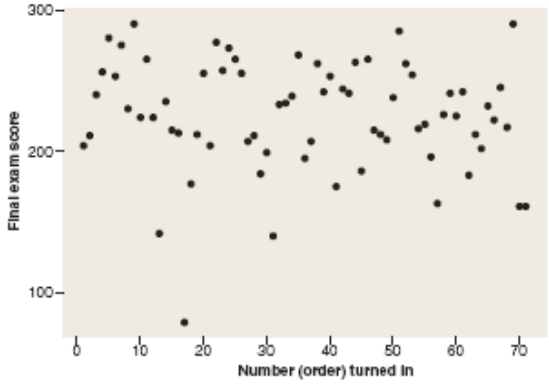
- **Correlation *r*:** tells direction and strength of linear relation between 2 quantitative variables
  - **Direction:** *r* is
    - positive for positive relationship
    - negative for negative relationship
    - zero for no relationship
  - **Strength:** *r* is between -1 and +1; it is
    - close to 1 in absolute value for strong relationship
    - close to 0 in absolute value for weak relationship
    - close to 0.5 in absolute value for moderate relationship

Example: *Extreme Values of Correlation*

- **Background**: Scatterplots show relationships…
  - (left) Price per kilogram vs. price per pound for groceries
  - (middle) Used cars' age vs. year made
  - (right) Students' final exam score vs. order handed in



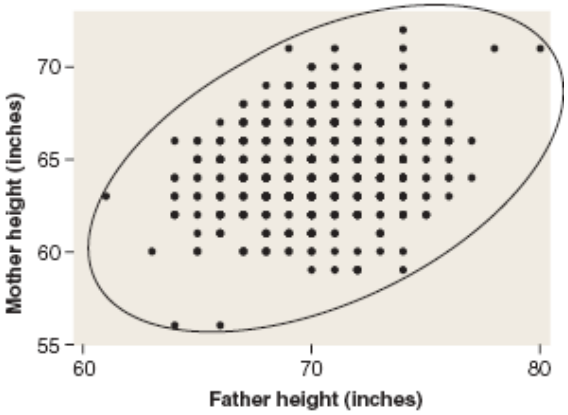


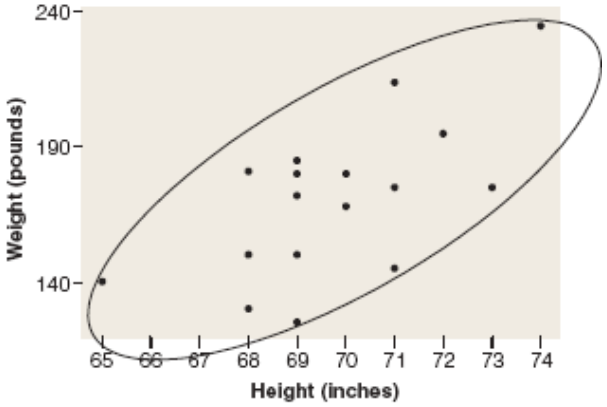


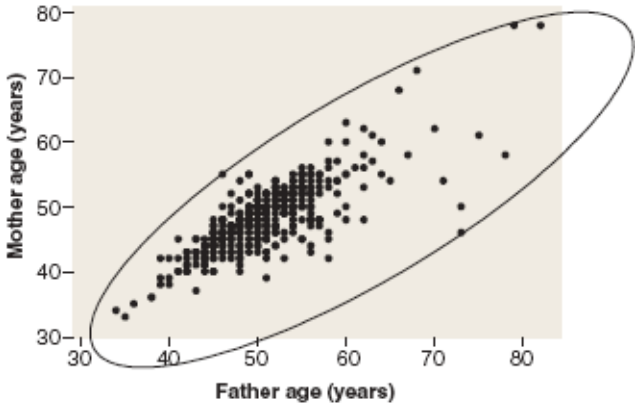
- **Question:** Correlations (scrambled) are -1, 0, +1. Which goes with each scatterplot?
- **Response:** left $r =$ _____ middle $r =$ _____ right $r =$ _____

Example: *Relative Strengths*

- **Background**: Scatterplots display:
  - mothers' ht. vs. fathers' ht. (left)
  - males' wt. vs. ht. (middle)
  - mothers' age vs. fathers' age (right):



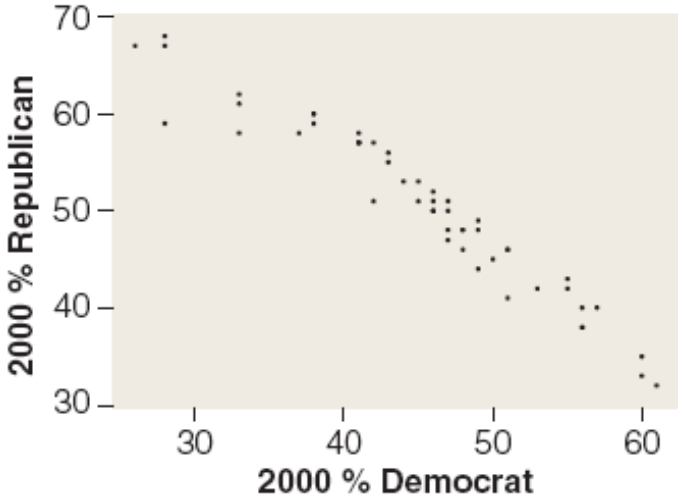




- **Question:** Which graphs go with which correlation: $r = 0.23$, $r = 0.78$, $r = 0.65$?
- **Response:** left $r =$ _____ middle $r =$ _____ right $r =$ _____

Example: *Imperfect Relationships*

- **Background**: For 50 states, % voting Republican vs. % Democrat in 2000 presidential election had $r = -0.96$.



- **Questions:** Why should we expect the relationship to be negative? Why is it imperfect?
- **Responses:**
  - Negative:
  - Imperfect:

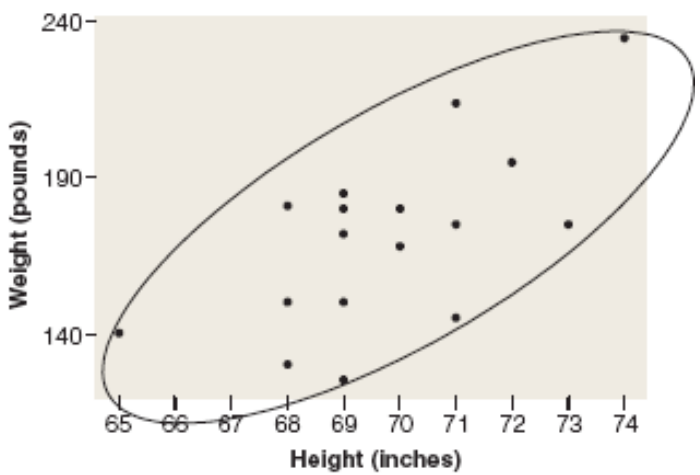
Practice: 5.37b p.193

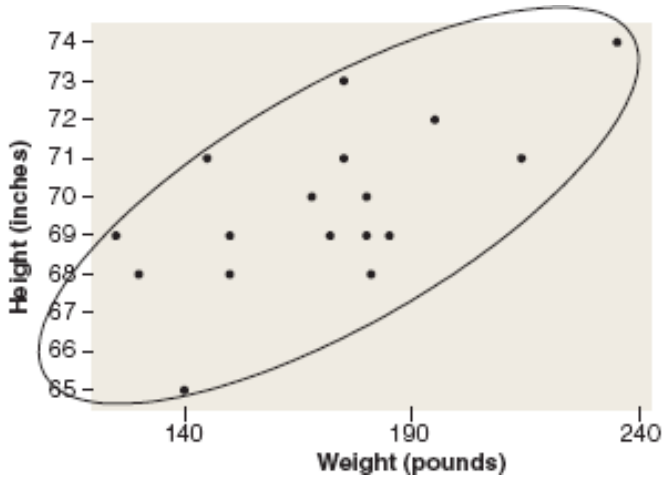
More about Correlation r

- Tells direction and strength of linear relation between 2 quantitative variables
  - A strong curved relationship may have *r* close to 0
  - Correlation not appropriate for categorical data
- Unaffected by roles explanatory/response
- Unaffected by change of units
- Overstates strength if based on averages

Example: *Correlation when Roles are Switched*

- **Background**: Male students' wt vs ht (left) or ht vs wt (right):



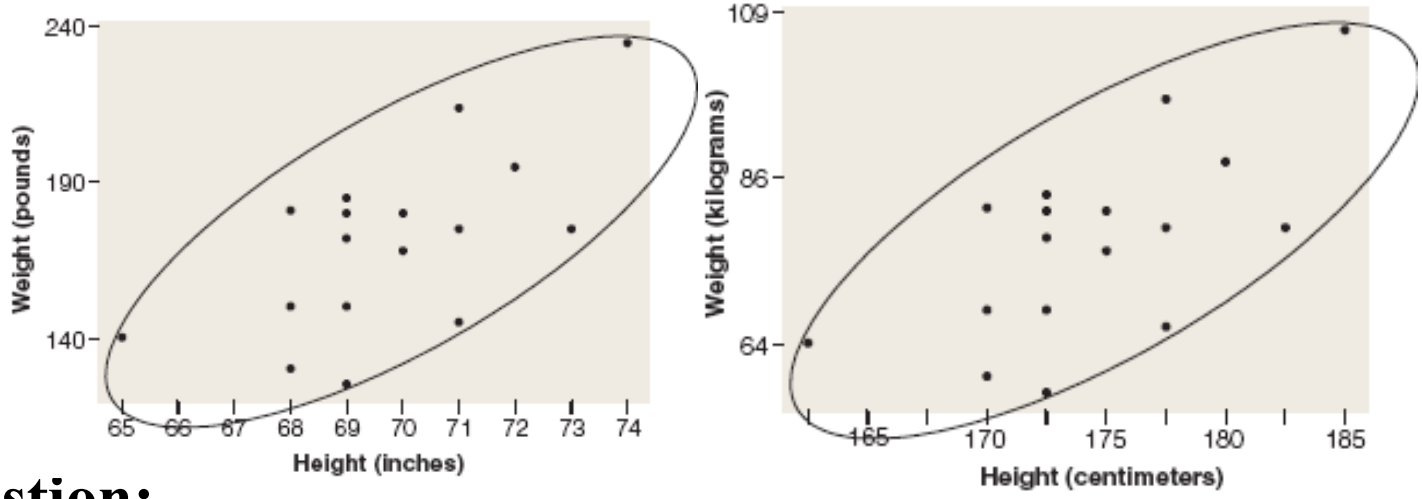


- **Questions:**
  - How do directions and strengths compare, left vs. right?
  - How do correlations *r* compare, left vs. right?
- **Responses:**
  -
  -

Example: *Correlation when Units are Changed*

- **Background**: For male students plot…

Left: wt (lbs) vs. ht (in) or **Right:** wt (kg) vs. ht (cm)



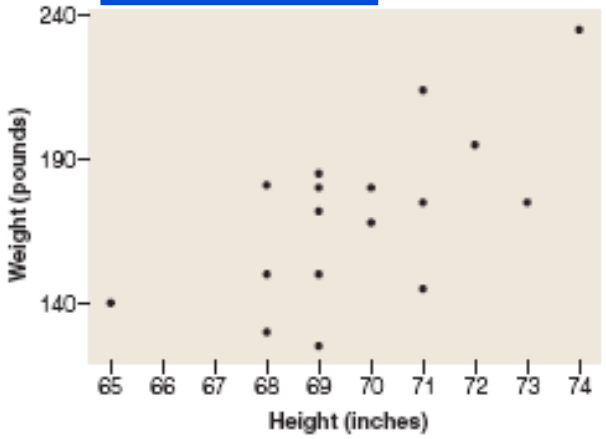
- **Question:**
  - How do directions, strengths, and *r* compare, left vs. right?
- **Response:**

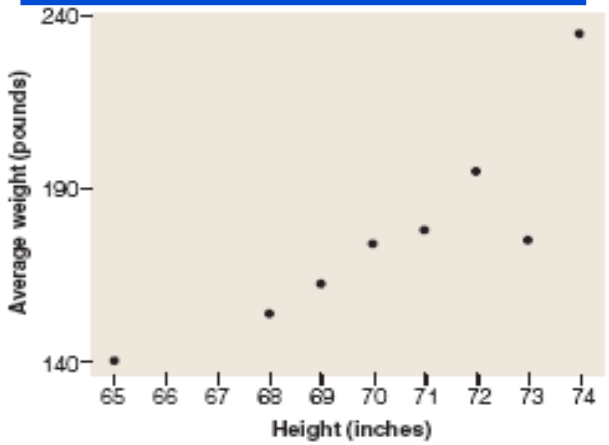
Example: *Correlation Based on Averages*

- **Background**: For male students plot…

Ht	65	68	69	70	71	72	73	74
Wt	140	130 150 181	125 150 172 180 185	168 180	145 175 214	195	175	235
AvWt	140	153.7	162.4	174.0	178.0	195	175	235

Left: wt. vs. ht. or **Right:** *average* wt. vs. ht.





- **Question:** Which one has r =+0.87? (other r =+0.65)
- **Response:** Plot on _____ has r = +0.87 (stronger).

Least Squares Regression Line

If form appears linear, then we picture points clustered around a straight line.

- **Questions** (Rhetorical):

1. Is there only one "best" line?
2. If so, how can we find it?
3. If found, how can we use it?

- **Responses:** (in reverse order)

3. If found, can use line to make predictions.

Least Squares Regression Line

- **Response:**

3. If found, can use line to make predictions.

Write equation of line $\hat{y} = b_0 + b_1 x$:

- Explanatory value is x
- Predicted response is $\hat{y}$
- y-intercept is b_0
- Slope is b_1

and use the line to predict a response for any given explanatory value.

Least Squares Regression Line

If form appears linear, then we picture points clustered around a straight line.

- **Questions:**

1. Is there only one "best" line?
2. If so, how can we find it?
3. If found, how can we use it? *Predictions*

- **Response:**

2. Find line that makes best predictions.

Least Squares Regression Line

- **Response:**

2. Find line that makes best predictions:

Minimize sum of squared *residuals* (prediction errors). Resulting line called **least squares line** or **regression line.**

A Closer Look: *The mathematician Sir Francis Galton called it the "regression" line because of the "regression to mediocrity" seen in any imperfect relationship: besides responding to x, we see y tending towards its average value.*

Least Squares Regression Line

If form appears linear, then we picture points clustered around a straight line.

- **Questions:**

1. Is there only one "best" line?
2. If so, how can we find it? *Minimize errors*
3. If found, how can we use it? *Predictions*

- **Response:**

1. Methods of calculus→unique "best" line

Least Squares Regression Line

If form appears linear, then we picture points clustered around a straight line.

- **Questions:**

1. Is there only one "best" line?
2. If so, how can we find it?
3. If found, how can we use it?

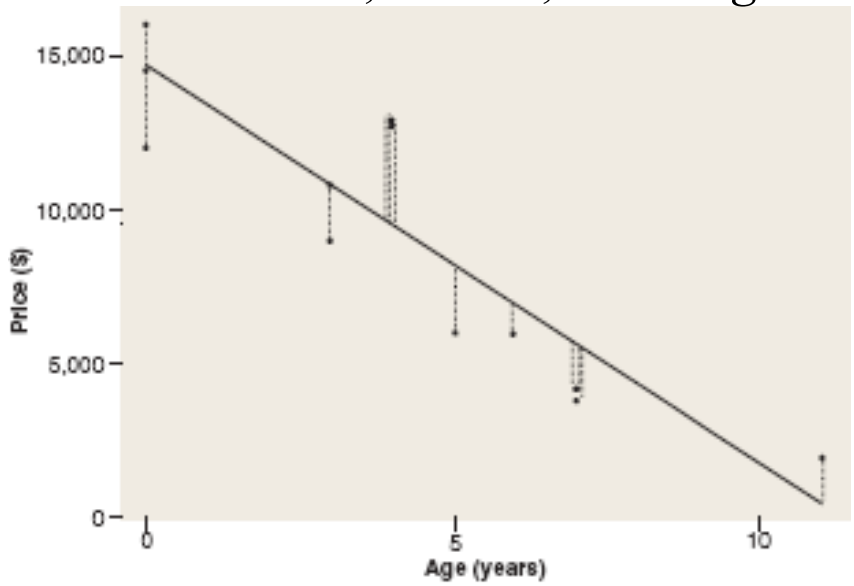
- **Response:**

1. "Best" line has $b_1 = r\frac{s_y}{s_x}$ $b_0 = \bar{y} - b_1\bar{x}$

Example: *Least Squares Regression Line*

- **Background**: Car-buyer wants to know if $4,000 is a fair price for an 8-yr-old Grand Am; uses software to regress price on age for 14 used Grand Am's:

$$Price = 14{,}686.3 - 1{,}287.64 Age$$



- **Question:** How can she use the line?
- **Response:** Predict for x=8, $\hat{y}$

Lecture Summary

(Quantitative Relationships; Correlation)

- Display with scatterplot
- Summarize with form, direction, strength
- Correlation r tells direction and strength
- Properties of *r*
  - Unaffected by explanatory/response roles
  - Unaffected by change of units
  - Overstates strength if based on averages
- Least squares regression line for predictions