

Section 9.1 - Correlation

1. Look at the relationship between variables. Observe that statistical relationships are overall tendencies, and they allow individual exceptions.
2. Explanatory and Response Variables
 - **Explanatory (independent) variable** - explains or influences changes in a response variable
 - **Response (dependent) variable** - measures an outcome of a study
 - When applicable, put explanatory variable on horizontal axis, response variable on vertical axis
3. Interpreting Scatterplots
 - (a) Look for an overall pattern and for striking deviations from that pattern.
 - (b) Describe the pattern by form, direction, and strength of the relationship.
 - **Form:** Look for clusters and for the “shape” (i.e. curved/linear/nothing/other)
 - **Direction:** Is it positively associated, negatively associated, or neither?
 - **Strength:** Determined by how closely the points follow a clear form.
 - (c) Locate outliers.
4. Correlation (**LINEAR** relationships) - Measures the direction and strength of the linear relationship between two quantitative variables. (Denoted by r)
 - Positive r indicates positive association; Negative r indicates negative association
 - $-1 \leq r \leq 1$
 - $|r|$ closer to 1 \Rightarrow stronger relationship
 - Does not change if we change units
 - Ignores distinction between explanatory and response variables
 - Strongly affected by a few outliers
 - If \bar{x}, \bar{y} and s_x, s_y are the mean and standard deviation for the x - and y - values for data on n individuals, then

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right).$$

- An average of the products of the standardized x -values and standardized y -values for n individuals.

5. Example: **Sports car gas mileage**

The following table gives the city and highway gas mileages for two-seater cars. Make a scatterplot that shows the relationship between city and highway mileage, using city mileage as the explanatory variable. Describe the overall pattern. Does the outlier (the Honda Insight) extend the pattern of the other cars or is it far from the line they form?

Model	City	Highway	Model	City	Highway
Acura NSX	17	24	Honda Insight	57	56
Audi TT Quattro	20	28	Honda S2000	20	20
Audi TT Roadster	22	31	Lamborghini Murcielago	9	9
BMW M Coupe	17	25	Mazda Miatta	22	22
BMW Z3 Coupe	19	27	Mercedes-Benz SL500	16	16
BMW Z3 Roadster	20	27	Mercedes-Benz SL600	13	13
BMW Z8	13	21	Mercedes-Benz SLK230	23	23
Chevrolet Corvetter	18	25	Mercedes-Benz SLK320	20	20
Chrysler Prowler	18	23	Porsche 911 GT2	15	15
Ferrari 360 Modena	11	16	Porsche Boxster	19	19
Ford Thunderbird	17	23	Toyota MR2	25	25

6. Correlation and Causation:

- (a) Is there a direct cause-and-effect relationship between the variables?
- (b) Is there a reverse cause-and-effect relationship between the variables?
- (c) Is it possible that the relationship between the variables can be caused by a third variable or a combination of several variables?
- (d) Is it possible that the relationship between two variables may be coincidence?