7 — Correlation

7.1 Scatterplots

A scatterplot shows the relationship between two sets of data. Each pair of data points is represented as a single point on the plane. The more linear our set of points are the stronger the relationship between the two data sets is.

7.1.1 Relationships in Data

Definition 7.1 — Correlation coefficient (Pearson's r). The Correlation coefficient, commonly referred to as Pearson's r, describes the strength of the relationship between two data sets. The closer $|r|$ is to 1 the more linear(stronger) our relationship. The closer r is to zero the more scattered(weaker) our relationship. To compute Pearson's r you can you can use the formula:

$$r = \frac{Covariance(x,y)}{S_x \cdot S_y}$$

On a Google Docs spreadsheet we can do

```
=Pearson(start cell for variable x : end cell for variable x,
         start cell for variable y : end cell for variable y)
```

Definition 7.2 — Coefficient of Determination(r^2). The coefficient of determination is the percentage of variation in the dependent variable (y) that can be explained by variation in the independent variable (x)

7.2 Practice Problems

Problem 7.1 A researcher wants to investigate the relationship between outside temperature and the number of reported acts of violence. For this investigation, what is the predictor (x) variable and what is the outcome (y) variable?

Problem 7.2 Given a correlation coefficient of -.95, what direction is the relationship and how do we know this? What is the strength of this relationship and how do we know this? In terms of strength and relationship, how does this correlation coefficient differ from one that is .95?

Problem 7.3 What does it mean if we have a coefficient of determination = .55?

Problem 7.4 If a researcher found that there was a strong positive correlation between outside temperature and the number of reported acts of violence, does this mean that an increase or decrease in temperature causes an increase or decrease in the number of reported acts of violence? Why or why not?