4.4 Cause and Effect

Data analysis involves much more than fitting a line or a curve to a set of data points.

Once you determine that there is a correlation, it is also important to consider how and why such a correlation exists – we need understand that there is a difference between *correlation* and *causality*.

CAUSAL RELATIONSHIPS

X causes
Y

1. **Direct Cause and Effect**
   a change in the independent variable causes a change
   in the dependent variable
   Ex. when driving, as km ↑, fuel ↓

Z causes
X & Y

2. **Common Cause**
   an external 3rd variable causes changes in the other two.
   Ex. As ice cream sales ↑, # of drowning incidents ↑.
   (common cause : hot weather)

Y causes X

3. **Reverse Cause and Effect**
   The dependent variable causes a change in the independent
   Ex. As amount of coffee consumed ↑, the amount of
   anxiety ↑ as well

4. **Accidental Relationships**

A correlation exists w/o any causal relationship

Ex. As amount of white paper used ↑, number of dogs adopted goes down.

5. **Presumed Relationships**

A relationship can make logical sense, but no common cause or cause & effect is apparent.

Ex. ~ one's self esteem ↑, their vocabulary level does as well

1

Example 1 – The scatterplot and line of best fit to the right show the relationships between the number of successful free throws made out of 10 attempts and the number of hours spent practicing for members of the Markville basketball team.
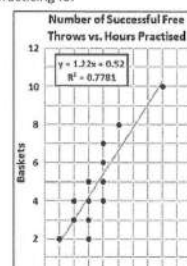
a) Characterize this correlation in regard to cause and effect.

$r^2 = 0.7781 \rightarrow r = 0.88$, strong positive

This is a cause & effect relationship b/c traditionally, more hours of practice leads to mastery of a skill.
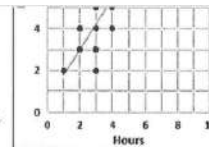
b) Interpret the line of best fit.

$b = 0.52$

exp...

Number of Successful Free Throws vs. Hours Practised

$y = 1.32x + 0.52$
$R^2 = 0.7781$

Baskets

Page 1 / 2

b) Interpret the line of best fit.

$b = 0.52 \rightarrow$ @ 0 hrs of practice, you can expect to make 0.52 shots

$a = 1.22 \rightarrow$ for every additional hr. of practice, you can make 1.22 more shots

Hours

c) Discuss any limitations of this linear model.

suggests that with continuous practice, you can exceed 10 baskets
(not possible).

Example 2 – Classify the relationship and justify your choice in each case.

a) A patient's stress level is negatively correlated with the amount of exercise performed.

reverse or cause & effect
→ exercise can reduce stress
→ stress can mean no time for exercise

b) Student math scores are positively correlated with English scores.

common cause → affinity for subjects
work ethic

c) Pancake sales are negatively correlated with amount of rainfall.

accidental → no clear cause

d) Job interview success rates are positively correlated with number of years a person has been married.

presumed → maturity
→ job experience
→ communication skills

2