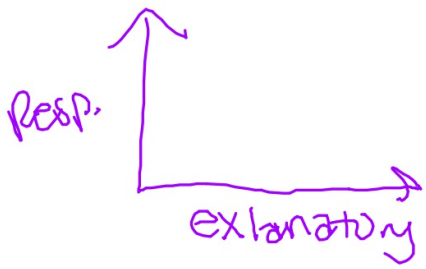
Chapter 3.1

Scatterplots and correlation

Why do we use scatterplots?

- Used to investigate relationships between **quantitative** variables.
- Types of Variables
  - Response Variable (Y)
  - Explanatory Variable (X)

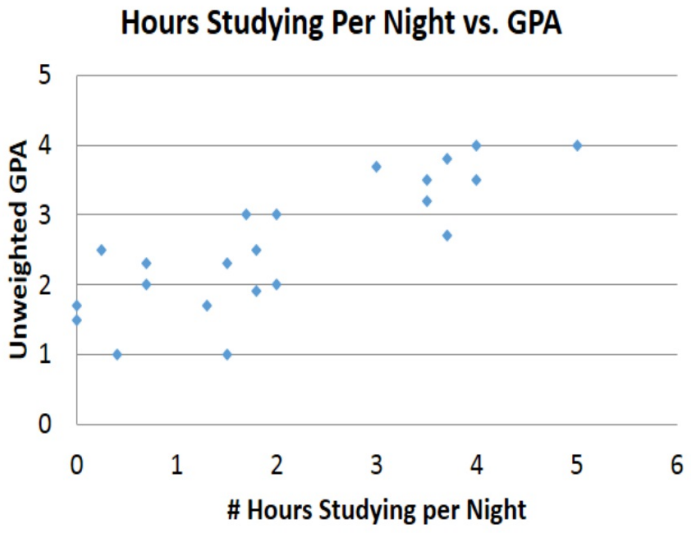


Interpreting Scatterplots

1. Direction : positive, negative
2. Form : linear, (nonlinear)
3. Strength: weak, moderate, strong
4. Outliers/Influential Points: Points that fall outside the pattern or are very far away from the other data points.

InterpretingScatterplots

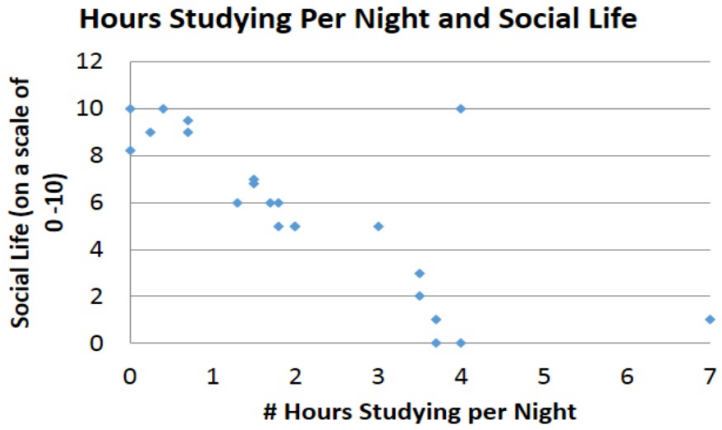
- Direction: Positive Association



- Form: Linear
- Strength: Moderate
- Outliers: None

InterpretingScatterplots

- Direction: Negative



Form: Linear

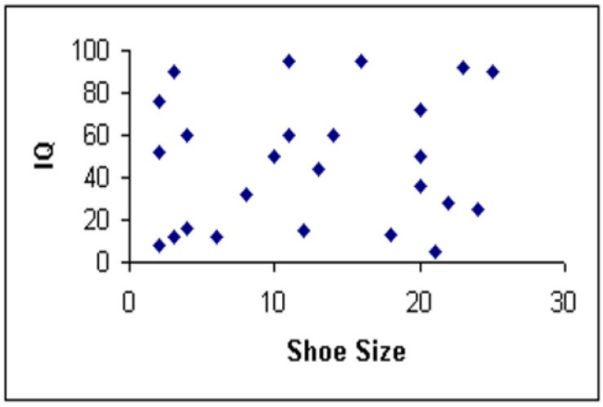
Strength: Moderately Strong

Outliers: Two outliers, (4,10) (7, 1)

InterpretingScatterplots

- Direction:

NO ASSOCIATION!!!



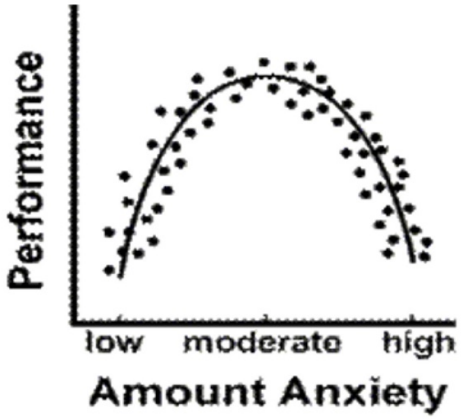
Form:

Strength:

Outliers:

InterpretingScatterplots

- Direction:



Form: Nonlinear (quadratic)

Strength: STRONG

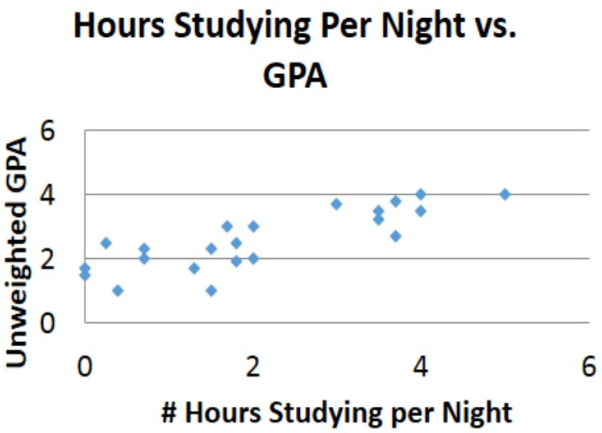
Outliers: None

Interpret IN CONTEXT

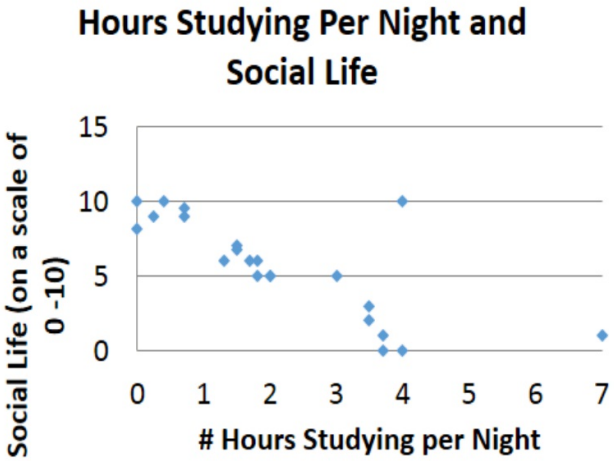
strength direct. form

- The scatterplot shows a (_____ _____ _____ association) between __(the variables)_____. There are no outliers/There is/are (an) outlier (describe outlier).

- The scatterplot shows a (____ ____ ____ association) between __(the variables)_____. There are no outliers/There is/are (an) outlier (describe outlier).



- The scatterplot shows a (_____ _____ association) between __(the variables)_____. There are no outliers/There is/are (an) outlier (describe outlier). The scatterplot appears _____.



Numbers!

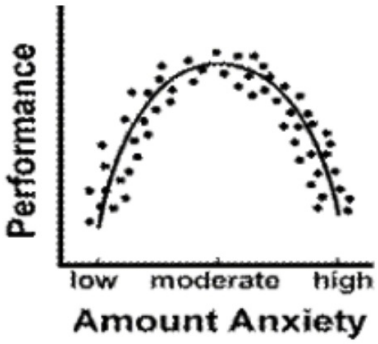
- If we are given the actual data, we can quantify the relationship between two variables.
- In addition to describing the association (the "visual" pattern), we can calculate the **correlation coefficient** (the numerical value that represents the association) r

Pearson's CorrelationCoefficient

- Karl Pearson (1857-1936) was an English lawyer, mathematician and statistician.
- He spent most of his career applying statistics in the field of biology.
- His contributions to the field of statistics include the correlation coefficient and the chi squared test (coming up later).

Pearson's Correlation Coefficient

- **Strength** of the **linear** relationship.
- *r*: Can take all values between -1 and 1 inclusive
  - r =-1 : perfect negative corr.
  - r = 0 : no correlation
  - r = +1 : perfect positive corr
  - r = 1.2 ← not possible ... oops
- A **perfect correlation:** all plotted points lie on a straight line.



0 – 0.25	0.25 – 0.5	0.5 – 0.75	0.75 - 1
very weak	weak	moderate	strong

Pearson's Correlation Coefficient

- $r = \frac{s_{xy}}{s_x s_y}$ where s_{xy} = covariance, s_x *and* s_y= standard deviations of x and y.
- $s_{xy} = \frac{\Sigma xy}{n} - \frac{\Sigma x}{n} \cdot \frac{\Sigma y}{n}$
- $s_x = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$ $\quad s_y = \sqrt{\frac{\Sigma y^2}{n} - \bar{y}^2}$

Pearson's Correlation Coefficient

- *Just kidding ….you can use your calculator* :)
- Enter the data from p. 177 into list 1 and list 2

CAUTION

- Did you know that there is an almost perfect correlation between ice cream sales and drowning deaths in Santa Monica?

CAUTION

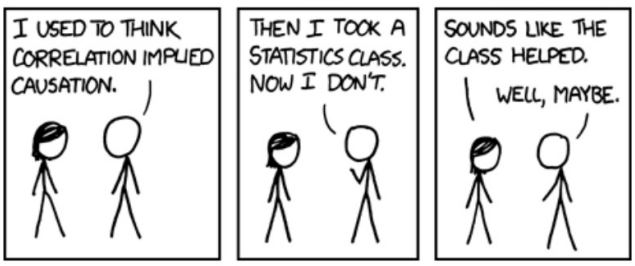
• Did you know that there is an almost perfect correlation between ice cream sales and drowning deaths in Santa Monica?

• **CORRELATION DOES NOT IMPLY CAUSATION!!!**

• In my class you may **not** use the word **cause** to describe a statistical relationship. Ever.

CAUTION

- Read p.191-193 **KNOWTHESE!!!!!!!**



Homework (after your activity)

3.5, 3.7, 3.8, 3.18, 3.20, 3.22 (just make a scatterplot in your cal, don't draw it) 3.23