3.2 Correlation

look at scatterplot:

form linear, curved, etc.

direction positive (increasing) or negative (decreasing)

strength perfect, nearly perfect, moderate, slight, etc.

"r" (correlation) = ***numerical measure*** of direction and strength of <u>linear</u> patterns

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

example: (64,7) (65,8.5) (63,8) ← 3 data points

x-bar = 64 y-bar = 7.83

s_x = .816497 s_y = .763763

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

(0)(-1.087) = 0

(1.225)(.8772) = 1.0746

(-1.225)(.2226) = -.273

\+

.8016

r = (1/2)(.8016) = .4008

Let's use an applet to make a scatterplot.

http://matti.usu.edu/nlvm/nav/frames_asid_144_g_3_t_5.html?open=activities&id=273

Let's examine some different r values visually.

http://espse.ed.psu.edu/statistics/Chapters/Chapter13/Chap13.html#Scatterplots

How well can you guess r from a scatterplot?

http://www.stat.uiuc.edu/courses/stat100//java/GCApplet/GCAppletFrame.html

Seven facts about correlation, r

1. Switching x & y won't change r.
2. x and y must be quantitative.
3. Units don't affect r & r has no units.
4. The sign r indicates direction of association.
5. $-1 \leq r \leq 1$ (near ± 1 means strong, near 0 means weak).
6. r measures linear relationships, not curved ones.
7. Correlation is not resistant-- outliers will affect r.

To report completely, find r, $\overline{x}$, $\overline{y}$, s_x, and s_y.