Day 21

Review A20

This converting back and forth to z-scores is a pain. (Use [A21pg.3/6d] as an example)

Can't we come up with a way to model the association that will work with the original units?

A Line! You do remember the equation of a line don't you? $y = mx + b$

A note of caution in statistics we write the equation of a line a bit differently: $\hat{y} = b_0 + b_1 x$

But what line ? Which is best?

$y - \hat{y}$; observed − predicted

Define residuals and show how they lead us to a definition of "line of best fit" (computer applets)

There must be a mathematical way to determine this "line of best fit"

To write the equation of a line we need the slope and a point ; y-intercepts

For example write the equation of a slope $\frac{1}{2}$, passing through $(4, -1)$

$y = \frac{1}{2}x + b$

$-1 = 2 + b$

$b = -3$

$y = \frac{1}{2}x - 3$

But what slope and point should we use to build our model?

If you suspect that what we know about correlation can lead us to the equation of the linear model you're headed in the right direction.

Last class we learned that by working with standardized data we could use the correlation coefficient to estimate the y value for a given x value. $\hat{z}_y = r z_x$

Looking at a scatterplot of the Standardized McDonald's data, is there any point that our line will definitely go through?