Correlation Example

x	y	z_x	z_y	$z_x\,z_y$	y values predicted by SD line	SD error
1	5	$-3/2$	$-1/2$	$3/4$	1	4
3	9	$-1/2$	$1/2$	$-1/4$	5	4
4	7	0	0	0	7	0
5	1	$1/2$	$-3/2$	$-3/4$	9	-8
7	13	$3/2$	$3/2$	$9/4$	13	0

- Compute the necessary statistics.

$$\begin{aligned}
\text{mean}(x) &= \frac{1}{5}\,(1+3+4+5+7) = 4 \\
\text{SD}(x) &= \sqrt{\frac{(1-4)^2+(3-4)^2+(4-4)^2+(5-4)^2+(7-4)^2}{5}} = \sqrt{\frac{9+1+0+1+9}{5}} = \sqrt{\frac{20}{5}} = \sqrt{4} \\
&= 2 \\
\text{mean}(y) &= \frac{1}{5}\,(5+9+7+1+13) = 7 \\
\text{SD}(y) &= \sqrt{\frac{(5-7)^2+(9-7)^2+(7-7)^2+(1-7)^2+(13-7)^2}{5}} = \sqrt{\frac{4+4+0+36+36}{5}} = \sqrt{\frac{80}{5}} \\
&= 4
\end{aligned}$$

- Convert the x values to standard units. For example,

$$x = 1 \quad \text{becomes} \quad z_x \frac{1-4}{2} = -3/2 \qquad \text{and} \qquad x = 3 \quad \text{becomes} \quad z_x \frac{3-4}{2} = -1/2$$

- Convert the y values to standard units: For example,

$$y = 5 \quad \text{becomes} \quad z_y = \frac{5-7}{4} = -1/2 \qquad \text{and} \qquad y = 9 \quad \text{becomes} \quad z_y \frac{9-7}{4} = 1/2$$

- Compute the products $z_x\,z_y$.
- Compute the correlation coefficient.

$$r = \frac{1}{5}\left(\frac{3}{4} - \frac{1}{4} + 0 - \frac{3}{4} + \frac{9}{4}\right) = \frac{1}{5}\,\frac{8}{4} = \frac{2}{5}$$

- Find the SD line.

$$\begin{aligned}
\big(y - \text{mean}(y)\big) &= (\text{sign }\ r)\,\frac{\text{SD}_y}{\text{SD}_x}\,\big(x - \text{mean}(x)\big) \\
(y-7) &= \frac{4}{2}\,(x-4) \\
y - 7 &= 2\,(x-4) \\
y &= 2\,x - 1
\end{aligned}$$

- Compute the y values predicted by the SD line. For example,

if $x = 1$, then $y = 2\cdot 1 - 1 = 2 - 1 = 1$ and if $x = 3$, then $y = 2\cdot 3 - 1 = 6 - 1 = 5$.

- Compute the SD errors (error $=$ y $-$ predicted value of y). For example,

if $x = 1$, then error $= 5 - 1 = 4$ and if $x = 3$, then error $= 9 - 5 = 4$.

- Compute the RMS size of the errors.

$$\text{RMS}_{\text{SD}} = \sqrt{(4^2+4^2+(-8)^2)/5} = \sqrt{(16+16+64)/5} = \sqrt{96/5} \approx 4.38.$$

Here is how to get `R` to do all this:

```
>source("http://www.adjoint-functors.net/SD.R")
>source("http://www.adjoint-functors.net/SDline.R")
> x <-c(1, 3, 4, 5, 7)                                   define the x values
> y <-c(5, 9, 7, 1, 13)                                  define the y values

> mean(x)                                                We can actually skip from here ...
> SD(x)
> mean(y)
> SD(y)
> zx <- zScore(x)
> zy <- zScore(y)
> r <- mean(zx*zy)
> SDslope <- sign(r)*SD(y)/SD(x)
> yIntercept <- mean(y) - SDslope*mean(x)                     ... to the next line

> SDline(x, y)                                           Read the slope, y intercept and r from this
> predictedBySD <- 2*x - 1
> SDerrors <- y - predictedBySD
> SDrms <- sqrt(mean(SDerrors^2))
```