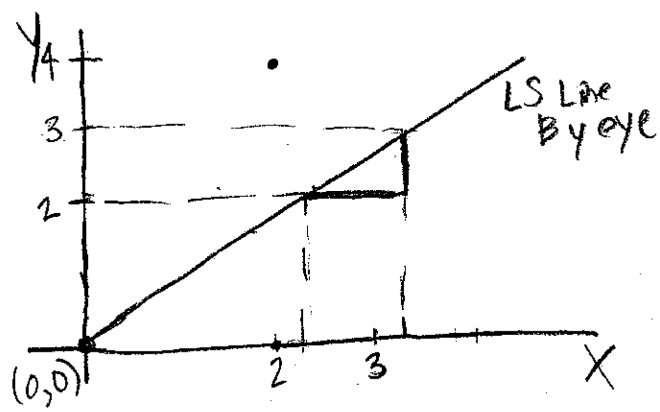
1. For the n = 4 data pairs below, determine the column means (do not evaluate square root of 3).

	x	y	x^2	y^2	xy
	0	0	0	0	0
	0	0	0	0	0
	2	0	4	0	0
	2	4	4	16	8
total	4	4	8	16	8
mean	1	1	2	4	2

2. a. Prepare an (x, y) scatterplot of the four points above and draw in the plot the least squares (i.e. regression) line as determined by eye.



run = 3.3 - 2.25 = 1.05

b. Give the numerical slope of your least squares line above.

$\text{Slope} = \frac{\text{rise}}{\text{run}}$

$\Rightarrow \frac{1}{1.05} = 0.95$

26

3. From your answers to (1) give the values below (show your work, do not reduce square root of 3).

a. sample sd s_x for x-scores = $\sqrt{\frac{n}{n-1}}\sqrt{\overline{x^2}-\bar{x}^2}$

$= \sqrt{\frac{4}{4-1}}\sqrt{2-1^2} = \sqrt{\frac{4}{3}}\sqrt{1} = \sqrt{\frac{4}{3}}$

b. sample sd s_y for y-scores = $\sqrt{\frac{n}{n-1}}\sqrt{\overline{y^2}-\bar{y}^2}$

$= \sqrt{\frac{4}{4-1}}\cdot\sqrt{4-1^2} = \sqrt{\frac{4}{3}}\cdot\sqrt{3}$

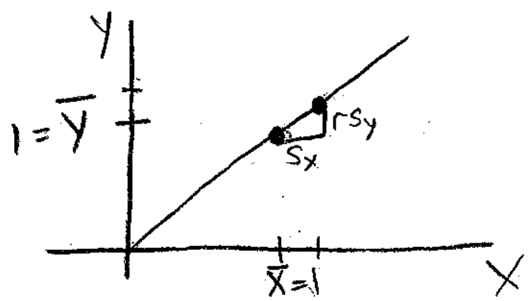
c. sample correlation $r = (\overline{xy} - \bar{x}\,\bar{y}) / \sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}$

$= [(2)-(1)(1)] / \sqrt{[(2)-(1)^2][(4)-(1)^2]} \Rightarrow \frac{1}{\sqrt{3}}$

d. slope of sample regression line = $r\, s_y / s_x$

$$r\,\frac{\cancel{\sqrt{\frac{n}{n-1}}}\sqrt{\overline{y^2}-\bar{y}^2}}{\cancel{\sqrt{\frac{n}{n-1}}}\sqrt{\overline{x^2}-\bar{x}^2}} = \frac{\sqrt{(4)-(1)^2}}{\sqrt{(2)-(1)^2}}\left(\frac{1}{\sqrt{3}}\right) = \left(\frac{\sqrt{3}}{\sqrt{1}}\right)\left(\frac{1}{\sqrt{3}}\right) = 1$$

e. sketch a plot of the line through the point x = mean x, y = mean y and having slope (d). It should agree with your line in (2).



4. a. (Keep in mind this example is artificial). A 95% confidence interval for the population mean μ_y of y-scores is given by

$$\bar{y} \pm 1.96\,\frac{s_y}{\sqrt{n}}$$

Evaluate this numerically but do not redue your answer.

$$1 \pm 1.96\left[\frac{\left(\sqrt{\frac{4}{3}}\sqrt{3}\right)}{\sqrt{4}}\right]$$

b. For larger with-replacement samples the claim is that [redacted] the 95% confidence interval calculated in the manner above has around 95% chance of doing what?

Covering μ_y.

c. A DIFFERENT 95% confidence interval for the population mean μ_y of y-scores incorporates both x and y scores and requires that we know the POPULATION mean μ_x of x-scores. It is given by

$$(\bar{y} - (\bar{x} - \mu_x)\, r\, s_y / s_x) \pm \sqrt{1 - r^2}\; 1.96\, \frac{s_y}{\sqrt{n}}$$

Express the above numerically if IS KNOWN THAT $\mu_x = 1.7$. Do not reduce.

$$1-(1-1.7)\left[\left(\tfrac{1}{\sqrt{3}}\right)\left(\sqrt{4/3}\,\sqrt{3}\right)/\left(\sqrt{4/3}\right)\right] \pm \sqrt{1-\left(\tfrac{1}{\sqrt{3}}\right)^2}\,(1.96)\,\frac{\left(\sqrt{4/3}\,\sqrt{3}\right)}{\sqrt{4}}$$

d. What advantage is claimed for confidence interval (c), which uses x-scores as well as y-scores, over confidence interval (a) which uses only y-scores?

Narrower for the same confidence level by a factor of $\sqrt{1-r^2}$