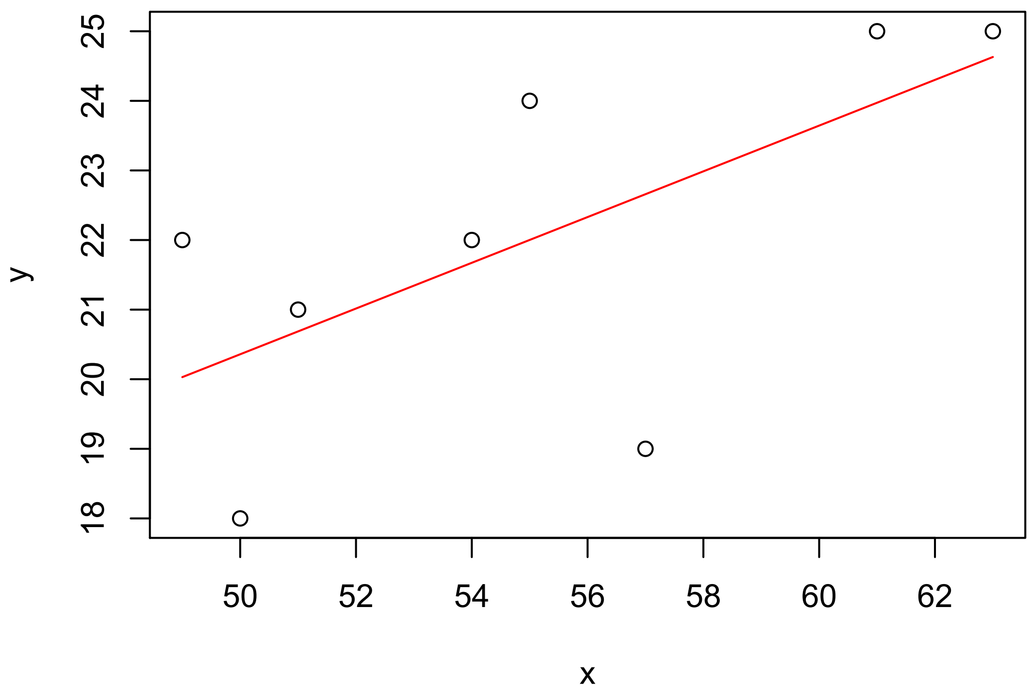
Psyc 3101, Homework 9

Imagine we have measured 8 subjects on two variables, *X* and *Y*. The data are below. You want to figure out how *X* and *Y* are related to each other. You'll do this first by calculating their correlation, and then by using each subject's *X* value to predict their *Y* value.

X	Y	$X-M_X$	$Y-M_Y$	z_X	z_Y	$z_X \cdot z_Y$	$\hat{z}_Y$	$\hat{Y}$	$Y-\hat{Y}$	$(Y-\hat{Y})^2$
54	22	-1	0	-.20	.00	.00	-.13	21.67	.33	.11
61	25	6	3	1.18	1.15	1.35	.75	23.97	1.03	1.06
49	22	-6	0	-1.18	.00	.00	-.75	20.03	1.97	3.89
57	19	2	-3	.39	-1.15	-.45	.25	22.66	-3.66	13.38
50	18	-5	-4	-.98	-1.53	1.50	-.63	20.36	-2.36	5.55
63	25	8	3	1.57	1.15	1.80	1.00	24.63	.37	.14
51	21	-4	-1	-.78	-.38	.30	-.50	20.69	.31	.10
55	24	0	2	.00	.76	.00	.00	22.00	2.00	4.00

1. Make a scatterplot of the data, by drawing a dot for each subject.



2. Write a guess for the correlation, and give a brief explanation for your guess.
I guess .6, because there's a positive relationship between x and y but it's far from perfect.

3. Calculate the means and standard deviations of *X* and *Y*.
$M_X = 55$, $\sigma_X = 5.10$, $M_Y = 22$, $\sigma_Y = 2.62$

4. Fill in the columns in the table for the deviations and z-scores.

5. Fill in the column for $z_X \cdot z_Y$.

6. Calculate the correlation. Compare to your guess in Question 2 (you don't need to write anything for this).
$r = \text{sum}(z_X \cdot z_Y)/(n-1) = .64$

The correlation is one measure of how well *X* and *Y* are related. Next you'll figure out how well you can use *X* to predict *Y*.

7. Use the correlation and each subject's z_X to predict their z_Y $(\hat{z}_Y)$ and enter these in the table.

$z_{Yhat} = r \cdot z_X$

8. Convert the predictions for z_Y to predictions for Y $(\hat{Y})$ and enter these in the table.

$Yhat = M_Y + r \cdot z_X \cdot s_Y$

9. Draw your prediction line on the scatterplot. An easy way to do this is to mark $(X, \hat{Y})$ for the smallest and largest values of *X*, and draw a line connecting them. For all the other subjects, $(X, \hat{Y})$ should fall on this line as well.

Finally, figure out how good the predictions are.

10. Fill in the errors and squared errors for the predictions.

11. Calculate the mean squared error of the predictions, MS_{Error}. This is the variance of *Y* that *X* cannot explain.

MS_{Error} = sum(Y-Yhat)^2/(n-1) = 4.03

12. Calculate the total variance of *Y*.
var(*Y*) = σ_Y^2 = 6.86

13. Subtract the residual variance from the total variance. This is the variance of *Y* that *X* can explain.
6.86 – 4.03 = 2.83

14. Divide the explained variance by the total variance. This is the proportion of the variance of *Y* that *X* can explain. Verify that the proportion of explained variance equals the square of your correlation.
2.83/6.86 = 41%
r^2 = .41 ✓