- **Within each of the plots below:**

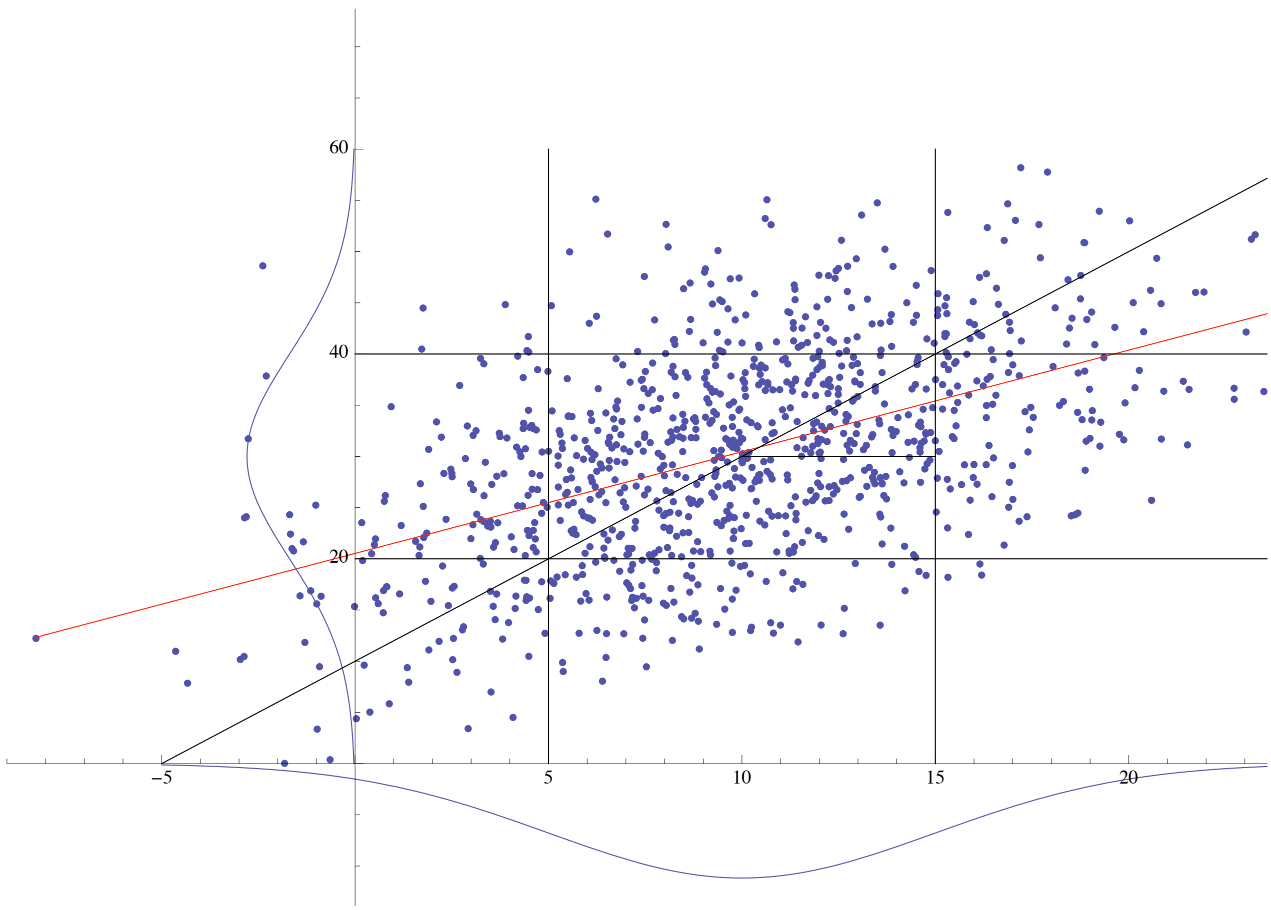
  Identify (theoretical) means and standard deviations of x, y.

  Identify the regression and naive lines.

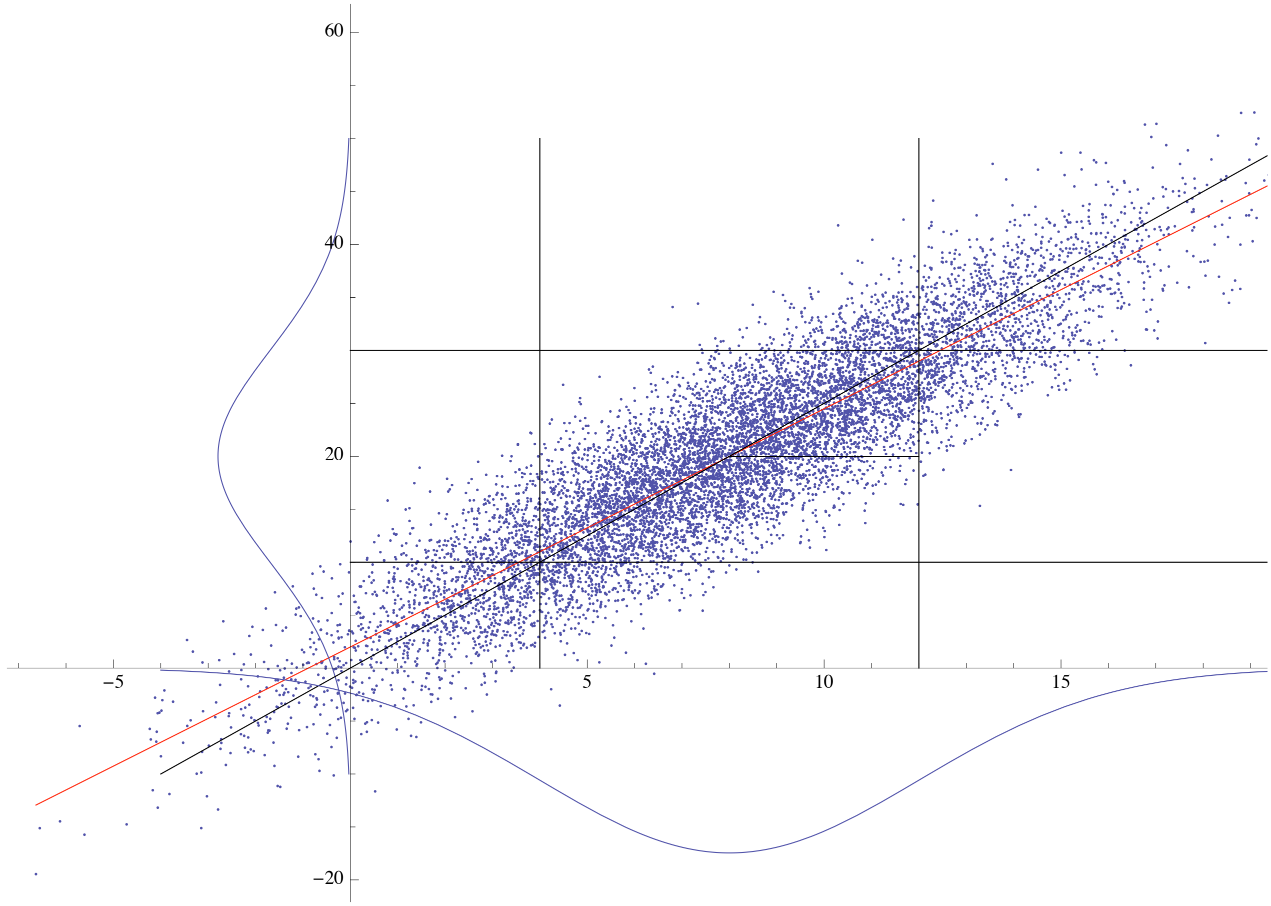
  Identify the correlation by noting the rise of the L. S. line relative to the naive line.

  Identify the sd of y for (any) fixed x and see from the given specimen it is $\sqrt{1 - r^2}\ s_y$.

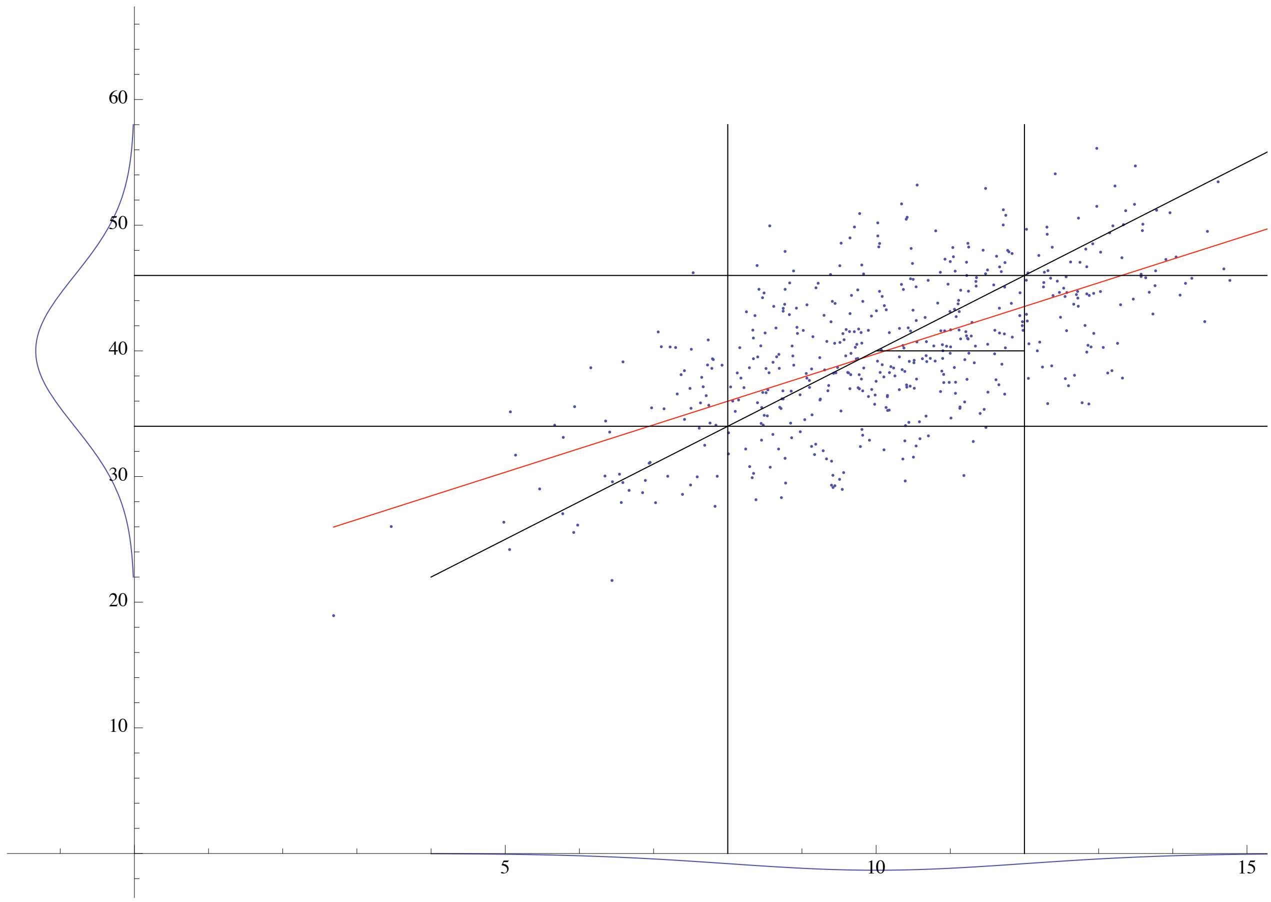
1. For this example the population means are 10, 30 the population sd are 5, 10. The correlation is 0.5.



2. Also, what is approximately the average y score for all of the points whose x is around 15?



3. Also, if x = 14 what are the mean and sd of the predicted value for y?



4. A sample of 200 homes is selected at random from a large community, with equal-probability and (effectively) replacement. For each home in the population

x = 2005 valuation y = 2010 valuation audit

It is known from the 2005 valuations that the x population mean is μ_x = 188,469. Each of the 200 sample homes is audited to determine the 2010 valuation audit. From the sample we find

$\bar{x}$ = 167,222 s_x = 74,873

$\bar{y}$ = 241,770 s_y = 156,224 r = 0.78

a. Ignoring x altogether and just using y-scores, give the 95% z-based CI for μ_y.

b. Determine the regression based 95% CI for μ_y given by

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

c. Compare the two CI (a), (b). How much wider is (a)? Notice that by exploiting regression in this way we've achieved a narrower 95% CI using the freely available x-scores associated with each sample home. Audits are by comparison very costly. So this beats using regular CI. We are required to know μ_x = 188,469 however.

d. Is $\overline{x}$ larger or smaller than the known μ_x? If it is smaller then since this r > 0 that would suggest that we should increase our estimate of μ_y upwards from $\overline{y}$. Do you follow this reasoning? Is it happening here?

5. For problem 33 page 190

x = time, y = calories (averages at bottom)

x	y	x^2	y^2	xy
21.4	472	457.96	222 784	10 100.8
30.8	498	948.64	248 004	15 338.4
37.7	465	1421.29	216 225	17 530.5
33.5	456	1122.25	207 936	15 276.
32.8	423	1075.84	178 929	13 874.4
39.5	437	1560.25	190 969	17 261.5
22.8	508	519.84	258 064	11 582.4
34.1	431	1162.81	185 761	14 697.1
33.9	479	1149.21	229 441	16 238.1
43.8	454	1918.44	206 116	19 885.2
42.4	450	1797.76	202 500	19 080.
43.1	410	1857.61	168 100	17 671.
29.2	504	852.64	254 016	14 716.8
31.3	437	979.69	190 969	13 678.1
28.6	489	817.96	239 121	13 985.4
32.9	436	1082.41	190 096	14 344.4
30.6	480	936.36	230 400	14 688.
35.1	439	1232.01	192 721	15 408.9
33.	444	1089.	197 136	14 652.
43.7	408	1909.69	166 464	17 829.6
_	_	_	_	_
34.01	456.	1194.58	208 788.	15 391.9

Let us suppose the data come from a random sample and the known population mean of time is 29 minutes.

a. Give the usual z-based 95% CI for the population mean of y ignoring x altogether.

b. Give the 95% z-based regression-based CI for the population mean of y.

c. How much narrower is (b) than (a)? Note: the sample size is rather small for the normal approximation to be confidently applied so regard this as a formal exercise of skills.