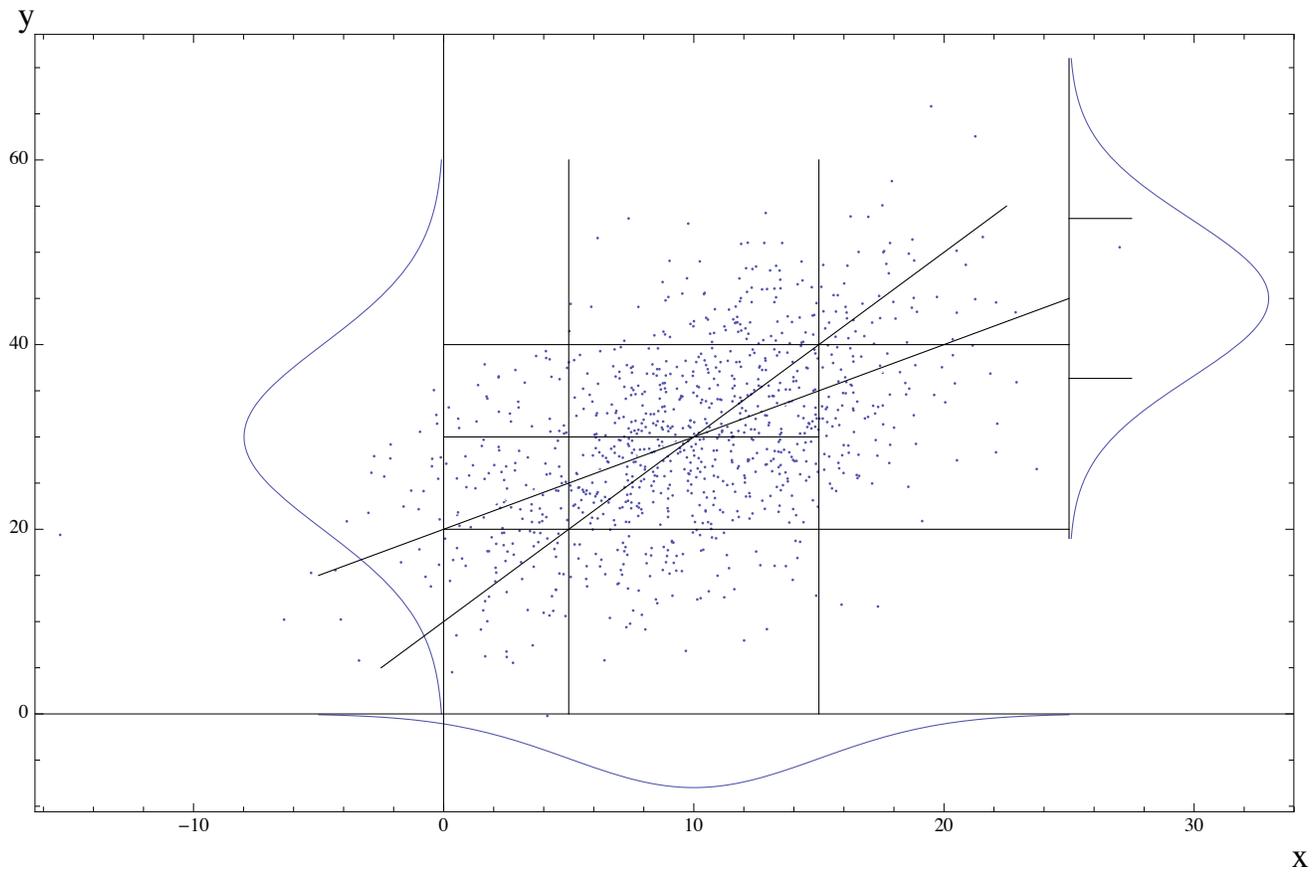


Questions 1 through 9 refer to the figure below. The distribution of  $(x, y)$  is 2-dimensional normal (bi-variate normal) vaguely represented by a large sample of points  $(x, y)$ . All curves plotted are for the population not for the sample.



1. Determine the population mean of  $x$  denoted  $\mu_x$ .
2. Determine the population standard deviation of  $x$  denoted  $\sigma_x$ .

3. Determine the *marginal probability* that  $y$  lies in the range  $[30, 40]$  denoted  $P(30 < Y < 40)$ .
4. Determine the *correlation*  $R$  between  $x$  and  $y$ .
5. **Using the naive line**, determine a *predicted value* of  $y$  for  $x = 25$ . Does it look as though this prediction is a good one?
6. **Using the regression line**, determine a *predicted value* of  $y$  for  $x = 25$ . Does it look as though this prediction is a good one?
7. Determine standard deviation of the bell curve in the upper right. It is the standard deviation of the *error of prediction* of  $y$  when **using the regression line** at  $x = 25$ . Do it by eye then confirm that your answer agrees with  $\sqrt{1 - R^2} \sigma_y$ .
8. Determine standard deviation of the *error of prediction* (**using the regression line**) at  $x = 12$ .
9. Determine a numerical *interval* for which the conditional probability of  $y$  being in the *interval* is around 0.68 if it is known that  $x = 25$ .

Questions 10 through 12 assume that IQ in a particular population is distributed as normal with mean 100 and standard deviation 15.

10. Determine the standard score  $z$  of a person having an IQ of 124.

11. Use the table of areas under the standard normal curve to determine the fraction of the population having  $\text{IQ} < 124$ . This fraction is the fraction of the standard normal to the left of the  $z$ -score of 124.

12. What is the IQ you need to have in order to surpass 83% of the population? This would place you at the 83rd percentile of IQ.

a. To find out, first find the 83rd percentile of  $z$  by entering 0.83 (or the closest value near it) in the body of the table of left-tail  $z$  areas, then read off the  $z$ -value.

b. After you find the 83rd percentile of  $z$  convert this to the 83rd percentile of IQ according to  $\text{IQ} = 100 + 15z$ .

So what is the 83rd percentile of IQ?

Questions 13 through 15 deal with algebraic properties of sample mean, sample standard deviation and correlation.

13. If sample mean of a list  $x$  is 15.2 what is the sample mean of list  $1.6x - 16$ ?

14. If sample standard deviation  $s_x$  of a list  $x$  is 3.4 what is sample standard deviation of list  $1.6x - 16$ ?

15. If correlation between  $(x, y)$  is 0.6 what is the correlation between  $(2x-4, 3y+2)$ ?

Questions 16 through 23 are about calculations of means, standard deviations, correlation in relation to the following data of  $(x, y)$  pairs (column means are recorded at the bottom):

$x$	$y$	$x^2$	$y^2$	$xy$
0	0	0	0	0
0	9	0	81	0
9	3	81	9	27
—	—	—	—	—
3.	4.	27.	30.	9.

16. Mean of  $x =$

17. Sample standard deviation of  $x =$

Note: Use a calculator program to get it then confirm in a separate calculation that it is equal to

$$s_x = \sqrt{\frac{n}{n-1}} \sqrt{\text{mean of } x \text{ squares} - \text{square of } x \text{ mean}}$$

18. Correlation R =

**Note: Use a calculator program and confirm it is equal to**

R =

mean of xy products – product of x mean by y mean /

$$\sqrt{\text{mean of } x \text{ squares} - \text{square of } x \text{ mean}} \sqrt{\text{mean of } y \text{ squares} - \text{square of } y \text{ mean}}$$

**Note: For the remaining problems we work with the naive and least squares line. Were the data *normal* then the least squares line would be (very close to) the regression line.**

19. Plot the three points  $(x, y)$  and locate the point of means  $(x \text{ mean}, y \text{ mean})$  then mark another point  $(x \text{ mean} + s_x, y \text{ mean} + s_y)$  and join these into a line. What is this line?

20. On the same plot as #19 plot the three points  $(x, y)$  and locate the point of means  $(x \text{ mean}, y \text{ mean})$  then mark another point  $(x \text{ mean} + s_x, y \text{ mean} + R s_y)$  and join these into a line. What is this line?

21. What is the value predicted for y if  $x = 0$ ?

22. Does this plot appear normal?

23. Is the least squares line for this data the same as the regression line?

Questions 24 through 28 concern T and normal based confidence intervals and the margin of error. Throughout, we suppose a random sample is selected with-replacement and with equal probability from a population (the distribution of the population  $x$  scores is not necessarily normal unless it is specifically assumed in a problem). Suppose the sample mean is 23.2, the sample standard deviation is 4.2, the sample size is 50, and the population size is 4800.

23a. Give the *point estimate* of the population mean based on this data.

24. Give the estimated margin of error for your estimate of #23a.

25. Give the 95% confidence interval for the population mean based on this data.

26. Your answer to #25 is an interval calculated from data without knowledge of the population mean. What is the *approximate probability* that such an interval covers the population mean? That means out of all possible intervals constructed from random samples.

27. If instead of sampling with replacement we sampled without replacement what would be your answer to #24?

28. If instead of sampling with replacement we sampled *without replacement* what would be your answer to #25?

29. If instead we had this same data but from a sample of only  $n = 6$  and if the population distribution is known to be close to normal what number would we use in place of the z-score 1.96 in #25? What is the applicable degrees of freedom?

30. For a sample of  $n = 6$  from a normal population give the 95% confidence interval for the population mean. Notice it is widened as compared with the  $n = 50$  case since  $n$  is smaller and therefore  $s / \sqrt{6}$  is larger than  $s / \sqrt{50}$  but also the applicable T score exceeds 1.96. THE T CONFIDENCE INTERVAL IS EXACT, MEANING IT ACHIEVES EXACTLY THE NOMINAL 0.95 COVERAGE PROBABILITY IF THE POPULATION DISTRIBUTION IS INDEED NORMAL, THE T-TABLE IS PERFECTLY ACCURATE AND ALL CALCULATIONS ARE INFINITE PRECISION.