



2. Lengths of pregnancies vary approximately according to a Normal distribution with mean 266 days and standard deviation 16 days.

- (a) Choose an Excel menu below (only one!), and fill it out to find the pregnancy length at the third quartile of the population.

**NORMDIST**

**X**  = number

**Mean**  = number


**Standard\_dev**  = number

**Cumulative**  = logical

=

Returns the normal cumulative distribution for the specified mean and standard deviation.

**X** is the value for which you want the distribution.

 Formula result =

**NORMINV**

**Probability**  = number


**Mean**  = number

**Standard\_dev**  = number

=

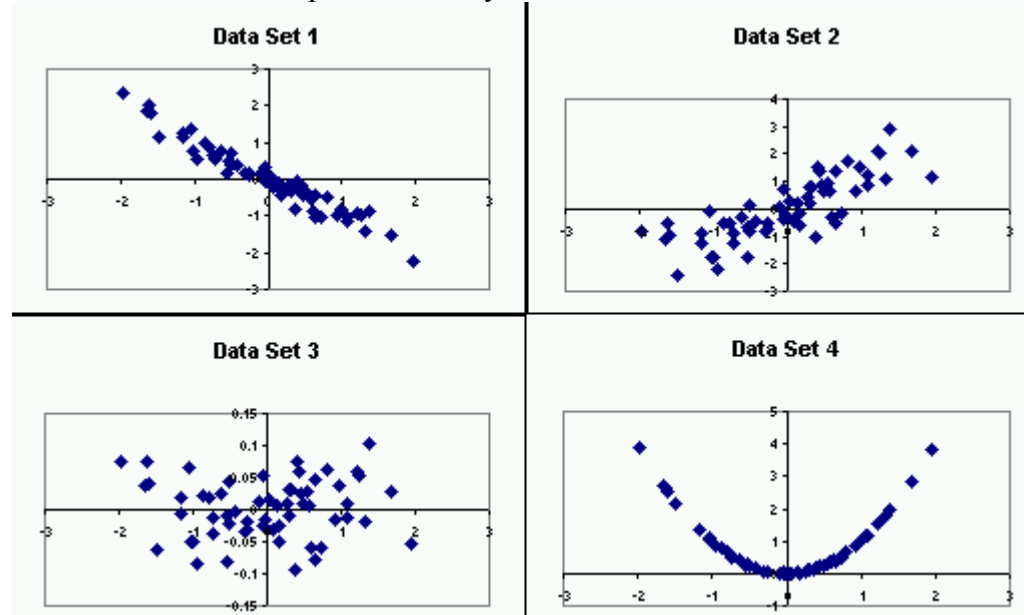
Returns the inverse of the normal cumulative distribution for the specified mean and standard deviation.

**Probability** is a probability corresponding to the normal distribution, a number between 0 and 1 inclusive.

 Formula result =

- (b) Write an Excel command to calculate the percent of pregnancies between 250 and 290 days
- (c) Write an Excel command to calculate the percent of pregnancies that are within 1.5 standard deviations of the mean.
- (d) Write an Excel command to calculate how long the longest 10% of pregnancies last.
- (e) Use the 68-95-99/7 rule to write an Excel command to calculate values between which the middle 95% of all pregnancies lie.

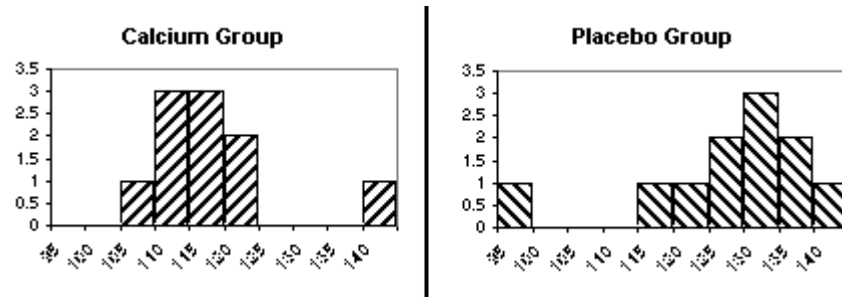
3. Here are scatterplots for 4 Toy Data Sets



Match the data sets to all statements which apply (matches can overlap, can be reused, and may not exist).

- |    |  |   |
|----|--|---|
| a. | Data Set 1<br>Data Set 2<br>Data Set 3<br>Data Set 4 | The variables are strongly associated with each other.<br>The variables are moderately associated with each other.<br>The variables are not associated with each other.               |
| b. | Data Set 1<br>Data Set 2<br>Data Set 3<br>Data Set 4 | The correlation is approximately $r = -0.95$<br>The correlation is approximately $r = 0$<br>The correlation is approximately $r = 0.8$<br>The correlation is approximately $r = 0.95$ |
| c. | Data Set 1<br>Data Set 2<br>Data Set 3<br>Data Set 4 | There are no obvious outliers<br>There is one serious outliers<br>There are two probably outliers<br>There are many outliers  |
| d. | Data Set 1<br>Data Set 2<br>Data Set 3<br>Data Set 4 | There is a positive linear relationship.<br>There is a negative linear relationship.<br>There is a curved relationship.<br>There is no apparent relationship.                         |

4. In a medical experiment, one group of men was given calcium, and another group was given a placebo. After some time their blood pressures were recorded and are shown in these two histograms.



Circle one of True, False or N.E.I. ("Not Enough Information") for each of the following:

- True False N.E.I The Calcium Group had higher blood pressures overall.
- True False N.E.I The Placebo Group population shows more variability.
- True False N.E.I There is a mild positive correlation between Cal. and Pla. Groups
- True False N.E.I The Placebo Group has an outlier to the left.
- True False N.E.I When the outlier is ignored, the Placebo Dist'n is left skewed.
- True False N.E.I When the outlier is ignored, the Cal. Dist'n is fairly symmetric.
- True False N.E.I The Calcium Distribution has 3 modes.
- True False N.E.I The median is larger for the Calcium Group than for the Controls.
- True False N.E.I The Inter Quartile Range is larger for the Calcium Group.
- True False N.E.I The range is larger for the Calcium Group.

5. To understand erosion, water was released on a test bed, at different flow rates, and the amount of eroded soil was measured.
- a. What is the response variable?
  
  
  
  
  
  
  
  
  
  
  - b. What is the explanatory variable?
  
  
  
  
  
  
  
  
  
  
  - c. If the x-data values are in Excel cells D4:D24, and the y data values are in the Excel cells E4:E24, write an Excel formula to calculate the y-intercept of the least squares regression line.
  
  
  
  
  
  
  
  
  
  
  - d. For data as in (c), write an Excel formula to calculate the slope of the least squares regression line.
  
  
  
  
  
  
  
  
  
  
  - e. If the y-intercept and slope from (c) are  $-3$  and  $2$  (respectively), write an Excel formula to calculate the predicted y value corresponding to a new x value of  $27$ .
  
  
  
  
  
  
  
  
  
  
  - f. If the x values range from  $5$  to  $15$ , is the prediction in (e) likely to be reasonably accurate? Explain why or why not in 20 words or less.