A sample size of 1000 or 200 million. That is the voting preferences proportion \( p \) that is of time. Therefore is the worst case and sample-size calculator turns out to be a confidence interval calculator you obtained.

The human brain wants to pass tests of randomness. Use in statistics software recommended. If it is you have to use to see if bias exists in

Form as possible, with it works best when it looks pretty and looks different gives a more gives a less precise of cost and efficiency to overcome any sampling from production

For example, if a effect the samples. to throw a die to

Decisions based

For assistance with sampling plan for quality

---

**scatter diagram**

Also called: scatter plot, X–Y graph

**Description**

The scatter diagram graphs pairs of numerical data, one variable on each axis, to look for a relationship between them. If the variables are correlated, the points will fall along a line or curve. The better the correlation, the tighter the points will hug the line.

**When to Use**

- When you have paired numerical data, and . . .
- When the dependent variable may have multiple values for each value of the independent variable, and . . .
- When trying to determine whether the two variables are related, such as:
  - When trying to identify potential root causes of problems, or . . .
  - After brainstorming causes and effects using a fishbone diagram, to determine objectively whether a particular cause and effect are related, or . . .
  - When determining whether two effects that appear to be related both occur with the same cause, or . . .
  - When testing for autocorrelation before constructing a control chart

**Procedure**

1. Collect pairs of data where a relationship is suspected.
2. Draw a graph with the independent variable on the horizontal axis and the dependent variable on the vertical axis. For each pair of data, put a dot or a symbol where the x-axis value intersects the y-axis value. If two dots fall together, put them side by side, touching, so that you can see both.
3. Look at the pattern of points to see if a relationship is obvious. If the data clearly form a line or a curve, you may stop. The variables are correlated. You may wish to use regression or correlation analysis now. Otherwise, complete steps 4 through 7.
4. Divide points on the graph into four quadrants. If there are \( X \) points on the graph,
   - Count \( X/2 \) points from top to bottom and draw a horizontal line.
   - Count \( X/2 \) points from left to right and draw a vertical line.
   If number of points is odd, draw the line through the middle point.
5. Count the points in each quadrant. Do not count points on a line.
6. Add the diagonally opposite quadrants. Find the smaller sum and the total of points in all quadrants.
   
   \[
   A = \text{points in upper left} + \text{points in lower right}
   \]
   \[
   B = \text{points in upper right} + \text{points in lower left}
   \]
   \[
   Q = \text{the smaller of } A \text{ and } B
   \]
   \[
   N = A + B
   \]
7. Look up the limit for \( N \) on the trend test table (Table 5.18).
   - If \( Q \) is less than the limit, the two variables are related.

<table>
<thead>
<tr>
<th>( N )</th>
<th>Limit</th>
<th>( N )</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–8</td>
<td>0</td>
<td>51–53</td>
<td>18</td>
</tr>
<tr>
<td>9–11</td>
<td>1</td>
<td>54–55</td>
<td>19</td>
</tr>
<tr>
<td>12–14</td>
<td>2</td>
<td>56–57</td>
<td>20</td>
</tr>
<tr>
<td>15–16</td>
<td>3</td>
<td>58–60</td>
<td>21</td>
</tr>
<tr>
<td>17–19</td>
<td>4</td>
<td>61–62</td>
<td>22</td>
</tr>
<tr>
<td>20–22</td>
<td>5</td>
<td>63–64</td>
<td>23</td>
</tr>
<tr>
<td>23–24</td>
<td>6</td>
<td>65–66</td>
<td>24</td>
</tr>
<tr>
<td>25–27</td>
<td>7</td>
<td>67–69</td>
<td>25</td>
</tr>
<tr>
<td>28–29</td>
<td>8</td>
<td>70–71</td>
<td>26</td>
</tr>
<tr>
<td>30–32</td>
<td>9</td>
<td>72–73</td>
<td>27</td>
</tr>
<tr>
<td>33–34</td>
<td>10</td>
<td>74–76</td>
<td>28</td>
</tr>
<tr>
<td>35–36</td>
<td>11</td>
<td>77–78</td>
<td>29</td>
</tr>
<tr>
<td>37–39</td>
<td>12</td>
<td>79–80</td>
<td>30</td>
</tr>
<tr>
<td>40–41</td>
<td>13</td>
<td>81–82</td>
<td>31</td>
</tr>
<tr>
<td>42–43</td>
<td>14</td>
<td>83–85</td>
<td>32</td>
</tr>
<tr>
<td>44–46</td>
<td>15</td>
<td>86–87</td>
<td>33</td>
</tr>
<tr>
<td>47–48</td>
<td>16</td>
<td>88–89</td>
<td>34</td>
</tr>
<tr>
<td>49–50</td>
<td>17</td>
<td>90</td>
<td>35</td>
</tr>
</tbody>
</table>
If $Q$ is greater than or equal to the limit, the pattern could have occurred from random chance.

**Example**

This example is part of the ZZ-400 improvement story on page 85. The ZZ-400 manufacturing team suspects a relationship between product purity (percent purity) and the amount of iron (measured in parts per million or ppm). Purity and iron are plotted against each other as a scatter diagram in Figure 5.172.

There are 24 data points. Median lines are drawn so that 12 points fall on each side for both percent purity and ppm iron.

To test for a relationship, they calculate:

$$A = \text{points in upper left} + \text{points in lower right} = 8 + 9 = 17$$

$$B = \text{points in upper right} + \text{points in lower left} = 4 + 3 = 7$$

$$Q = \text{the smaller of } A \text{ and } B = \text{the smaller of } 7 \text{ and } 17 = 7$$

$$N = A + B = 7 + 17 = 24$$

Then they look up the limit for $N$ on the trend test table. For $N = 24$, the limit is 6. $Q$ is greater than the limit. Therefore, the pattern could have occurred from random chance, and no relationship is demonstrated.
Considerations

- In what kind of situations might you use a scatter diagram? Here are some examples:
  - Variable A is the temperature of a reaction after 15 minutes. Variable B measures the color of the product. You suspect higher temperature makes the product darker. Plot temperature and color on a scatter diagram.
  - Variable A is the number of employees trained on new software, and variable B is the number of calls to the computer help line. You suspect that more training reduces the number of calls. Plot number of people trained versus number of calls.
  - To test for autocorrelation of a measurement being monitored on a control chart, plot this pair of variables: Variable A is the measurement at a given time. Variable B is the same measurement, but at the previous time. If the scatter diagram shows correlation, do another diagram where variable B is the measurement two times previously. Keep increasing the separation between the two times until the scatter diagram shows no correlation.
  - Even if the scatter diagram shows a relationship, do not assume that one variable caused the other. Both may be influenced by a third variable.
  - When the data are plotted, the more the diagram resembles a straight line, the stronger the relationship. See Figures 5.39 through 5.42, page 198 in the correlation analysis entry, for examples of the types of graphs you might see and their interpretations.
  - If a line is not clear, statistics (N and Q) determine whether there is reasonable certainty that a relationship exists. If the statistics say that no relationship exists, the pattern could have occurred by random chance.
  - If the scatter diagram shows no relationship between the variables, consider whether the data might be stratified. See stratification for more details.
  - If the diagram shows no relationship, consider whether the independent (x-axis) variable has been varied widely. Sometimes a relationship is not apparent because the data don’t cover a wide enough range.
  - Think creatively about how to use scatter diagrams to discover a root cause.
  - See graph for more information about graphing techniques. Also see the graph decision tree (Figure 5.68, page 216) in that section for guidance on when to use scatter diagrams and when other graphs may be more useful for your situation.
  - Drawing a scatter diagram is the first step in looking for a relationship between variables. See correlation analysis and regression analysis for statistical methods you can use.