

1.2: The Statistical Analysis Process

Statistics is the science of collecting, organizing, summarizing, and analyzing information to draw conclusions or answer questions. Additionally, statistics is about providing a measure of confidence in any conclusions. Statistics play a big role in our day to day decision making. Data can help us answer many questions:

- Students can use data to help pick a college that is a good fit for them
- Teachers use data to improve their teaching methods
- Medical researchers use data to learn if treatments actually work sufficiently
- Scientists use data to measure the effect humans have on the environment
- Car companies use data to determine how safe a vehicle is

Statistical analysis is the process of looking at a sample of data to learn something about a larger population that may be difficult to understand because of its size. It allows us to make generalizations about populations based on sample data. There are four steps:

1. Ask a question that can be answered by collecting data.
2. Decide what to measure and collect the data.
3. Summarize the data and analyze the data.
4. Draw a conclusion and communicate the results to your audience.

Step 1: Ask a question that can be answered by collecting data

Many people believe that personality, habits, likes and dislikes are affected by the time of year you were born. One such theory involves chronotypes. The term chronotype refers to a person's natural tendency to be most active and alert at certain times of day. People who are most active early in the day are labeled morning people, early birds, or larks. Those at their best in the evening are called night people or owls.

This chronotype theory states that there are three chronotypes and that each corresponds to four birth months.

Chronotype	Birth Months
Morning	May, June, July, August
Evening	January, February, November, December
None	March, April, September, October

We are going to use the four step statistical analysis process to try to answer this question: Is there a significant correlation between someone's birth month and their chronotype?

Step 2: Decide what to measure and then collect data

Instead of collecting data from all **individuals** of an entire group for the study (called the **population**), we could instead select a **sample** of the population. In order to investigate this question, we need to collect data from a random group of participants. We would need to know the _____ and _____ of each individual. We could have participants take an assessment to determine their chronotype. We would see what proportion of the participants had matching birth month and chronotype according to the chronotype theory (found in the table above).

Step 3: Summarize and analyze the data

In a group of 30 randomly selected adults, it was found that 11 had matching chronotype and birth month according to the provided chronotype theory. The proportion (percentage) of matches was _____. Is this proportion high enough to convince us that this theory applies to the entire population? In other words, can we infer the results to some group bigger than those in the sample?

Suppose this chronotype theory is false. Is it possible that a participant in the study could still select a chronotype that matched their birth month according to the theory? What fraction or proportion of the participants in the study do we expect to have matching birth month and chronotype according to the theory?

In order to analyze the data, we need to use probability to obtain strong enough evidence to support or reject a claim. If this chronotype theory is false, the birth month does not match the corresponding chronotype, there is a _____ chance that a student would select the chronotype predicted. How far above _____ would the proportion need to be in order to convince us that this chronotype theory is reasonable?

Chance variation is the type of differences we would naturally expect to see between many different samples. We will see what proportions in a sample are likely to occur just by chance by rolling a die. In a 6-sided die, each of the six outcomes are equally likely to occur. The event of rolling a 1 or a 2 should occur around 33% or exactly $\frac{1}{3}$ of the time. If you'd like to try this yourself, type in "roll dice" into google. Click the "roll" button and record your outcomes.

Die value	3	2	5	4	2	5	5	1	1	3	4	4	1	1	3	1	3	4	4	2
Resulted in 1 or 2	N	Y	N	N	Y	N	N	Y	Y	N	N	N	Y	Y	N	Y	N	N	N	Y

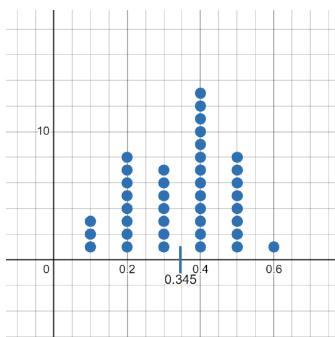
Proportion of die rolls resulting in 1 or 2: $\frac{8}{20} = 0.4 = \underline{\quad}\%$

Die value	6	4	5	2	6	1	1	5	5	1	2	5	2	2	1	4	1	1	4	5
Resulted in 1 or 2	N	N	N	Y	N	Y	Y	N	N	Y	Y	N	Y	Y	Y	N	Y	Y	N	N

Proportion of die rolls resulting in 1 or 2: $\underline{\quad} = \underline{\quad} = \underline{\quad}$

Note: A **proportion** is a number between 0 and 1. It represents a fraction or portion of a total. We usually write proportions as decimals or percents. To calculate the decimal, use a calculator to divide the numerator (top of the fraction) by the denominator (bottom of the fraction). For example, if you roll the die 25 times and 7 of those rolls resulted in a 1 or a 2, then the proportion as a fraction would be written as $\frac{7}{25}$ and you would enter $7 \div 25$ in your calculator to get 0.28. To change this to a percent, we multiply the decimal by 100 or move the decimal twice to the right and write a percent symbol. For example, $0.28 \cdot 100\% = 28\%$.

You can see from above that we did not get the same proportion in both times we tried this experiment, and the proportions we found were both close to $\frac{1}{3}$. We could repeat this experiment many times to see what outcomes are unusual. Below is a desmos graph of a dotplot of proportions from 40 repetitions of this experiment.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

This is an example of a **distribution** of data. We can now use it to make a decision about chronotype theory.

Step 4: Draw a conclusion and communicate the results

In the study, we found the proportion of matching chronotypes to birth month was around ____%. Since this proportion is likely to occur based on the distribution of data above, we can't generalize this result to the population because the proportion isn't high enough (or unusual enough) to convince us that the chronotype theory is true.

Identifying important information:

A group of researchers wondered if fewer than half of the women who visit a particular fertility clinic would want to choose the sex of their future child if it was possible. A total of 561 women responded to the survey. 229 of them said they wanted to choose the sex of their future child. The researchers analyzed the data and concluded that there was convincing evidence that fewer than half of the women who visit the clinic would choose the sex of a future child because, if, in reality, at least half of women who visit a fertility clinic would like to choose the sex of a future child, it would be very unusual to observe a percentage as low as 41% in a sample of 561 women.

1. What is the question being asked?
2. What is the way the data was collected and measured? What is the population? What is the sample?
3. Summarize and analyze the data:
4. What is the conclusion of the study?