# Questions Module 2.1

In the text box beside each characteristic, enter the letter of the most appropriate modeling type:

**a** for Nominal
**b** for Ordinal
**c** for Continuous

Scroll down if necessary and click the **Check My Answer** button. The correct answer will display below the button.

- [ ] wafer thickness
- [ ] pass/fail
- [ ] small, medium, large
- [ ] number of nonconformances
- [ ] defect type: pits, scratches, cracks
- [ ] part width
- [ ] percent impurity
- [ ] hair color

---

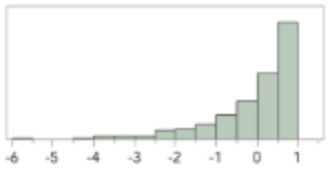
The correct answers from top to bottom are **c**, **a**, **b**, **c**, **a**, **c**, **c**, **a**. The pass/fail, defect type: pits, scratches, cracks, and hair color characteristics are Nominal. The small, medium, large characteristic is Ordinal. The wafer thickness, number of nonconformances, part width, and percent impurity characteristics are Continuous.

In the text box beside each of the four descriptions, enter the letter of the appropriate distribution shape. Scroll down if necessary and click the **Check My Answer** button. The correct answer will display below the button.

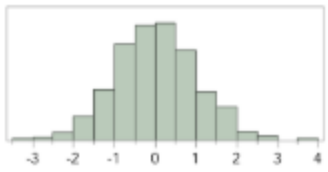
- [ ] bimodal a.
- [ ] right-skewed b.
- [ ] left-skewed c.
- [ ] normal d.

---

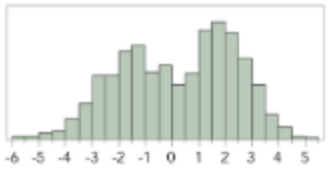
a.



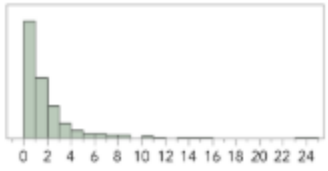

b.




c.




d




**Incorrect**.
The correct answers from top to bottom are **c**, **d**, **a**, **b**.

If a distribution is left-skewed, the median is most likely greater than the mean.

- *a.* True
- *b.* False

---

**Correct**.
For left-skewed data, the median is generally greater than the mean. For right-skewed data, the median is generally less than the mean.

Your score on a standardized test is in the 75$^{th}$ percentile. This means that 75% of people taking the test score higher than you do.

- *a.* True
- *b.* False

---

**Correct**.
At the 75$^{th}$ percentile, your score is higher than 75% of the people taking the same test.
Let's consider the historical data used in the White Polymer case study. Your team goal is to improve the yield of the polymer process. The crisis team data, which you used in the previous practice, were collected by another team. The KPI, **Yield**, should be closely related to the two primary output characteristics, **MFI** and **CI**.

What are some potential issues with using historical data in problem solving? Can you have confidence that the historical data will be useful in solving the problem?

Answer (there are many possible answers): It can be risky to use historical data, for many reasons. Here are a few:

- The process might have changed since the data were collected, so the data might not be relevant to the problem.
- The quality of the data might be questionable.
- The measurement systems might not be capable of accurately or precisely measuring the characteristics of interest.
- The data set might be missing important variables needed to solve the problem.

Unless you know the source of the data, can verify the quality of the data, and have verified that the measurement systems are capable, you must take caution when using historical data. Here are some potential next steps:

- Investigate the source of the historical data.
- Determine whether there have been recent measurement studies for the systems that measure the critical characteristics.
- Conduct measurement system studies.
- Develop a process map, a cause-and-effect diagram, and a data collection plan, and collect new data.
- Conduct a designed experiment.
-

You learn about measurement system studies and designed experiments in future modules.

If a data set contains an outlier, it should always be excluded from the analysis or deleted from the data set.

- *a.* True
- *b.* False

---

**Correct**.
The observation should be investigated. You might exclude the observation if there is an assignable cause.

The interquartile range contains the middle 50% of the observations in a distribution.

- *a.* True
- *b.* False

---

**Correct**.
The interquartile range is the middle 50% of the observations. It is the difference between the third quartile and the first quartile.

In the text box beside each of the four statistics, enter the letter of the appropriate definition.

| | |
|---|---|
| standard deviation | a. the spread of the middle 50% of the observations in a data set |
| range | b. a measure of the average distance between each observation and the mean |
| interquartile range | c. the average squared difference between each observation and the mean |
| variance | d. the difference between the largest and the smallest value |

**Incorrect**.
The correct answers from top to bottom are **b**, **d**, **a**, **c**. Note that, technically, the standard deviation is the square root of the average squared difference between each observation and the mean.

In the text box beside each of the three graph names, enter the letter of the appropriate description.

| | |
|---|---|
| run chart or line graph | a. plot showing the relationship between two continuous variables |
| comparative box plot | b. plot showing the behavior of a continuous characteristic over time |
| scatterplot | c. graph showing the distribution of a continuous variable for different levels of a categorical variable |

**Incorrect**.
The correct answers from top to bottom are **b**, **c**, **a**.

A mosaic plot is used to show the frequencies of one categorical variable.

- *a.* True
- *b.* False

**Incorrect**.
The correct answer is **b**. Mosaic plots are used to compare the frequencies, counts, or percents across levels of two categorical variables.

Bar charts can be used to graph more than one variable at a time.

- *a.* True
- *b.* False

**Correct**.
Like mosaic plots, bar charts (stacked, side-by-side, and so on) can be used to compare the frequencies, counts, or percents across levels of two categorical variables.

In the previous practice, you explored the relationships between some of the variables in the **VSSTeamData.jmp** data set. You focused on understanding the relationship between the KPI, **Yield**, and two output characteristics, **MFI** and **CI**.

You also started to explore the relationships between **MFI**, **CI**, and the input variables. You continue this exploration in future practices.

What can you learn from an exploratory analysis of **MFI**, **CI**, and the input variables?

Answer: There are many possible answers. Here are a few:

- Exploratory analysis can help you identify data quality issues, like missing values or outliers, that need to be addressed.
- Exploratory analysis can help you identify patterns in the data.
- Exploratory analysis can help you understand potential relationships between **MFI**, **CI** , and the input variables.
- Exploratory analysis can help you identify potential root causes of variation in **MFI** and **CI**.
- Exploratory analysis can guide you to the types of analyses that make sense, given the data and the variables.

Note that you learn more about exploratory data analysis and are introduced to tools for exploring many variables at a time, later in this module.