

Role-Type Classification

CO-4: Distinguish among different measurement scales, choose the appropriate descriptive and inferential statistical methods based on these distinctions, and interpret the results.

Video

Video: [Role-Type Classification](#) (Two Parts; 9:46 total time)

While it is fundamentally important to know how to describe the distribution of a single variable, most studies pose research questions that involve exploring the relationship between **two** (or more) variables. These research questions are investigated using a sample from the population of interest.

Reading: [Form a Research Question](#) (short)

Here are a few examples of such research questions with the two variables highlighted:

✓ EXAMPLES:

1. Is there a relationship between **gender** and **test scores** on a particular standardized test? Other ways of phrasing the same research question:
 - Is performance on the test related to gender?
 - Is there a gender effect on test scores?
 - Are there differences in test scores between males and females?
2. How is the **number of calories** in a hot dog related to (or affected by) the **type of hot dog** (beef, meat or poultry)? In other words, are there differences in the number of calories among the three types of hot dogs?
3. Is there a relationship between the **type of light** a baby sleeps with (no light, night-light, lamp) and whether or not the child develops **nearsightedness**?
4. Are the **smoking habits** of a person (yes, no) related to the person's **gender**?
5. How well can we predict a student's freshman year **GPA** from his/her **SAT score**?
6. What is the relationship between driver's **age** and sign legibility **distance** (the maximum distance at which the driver can read a sign)?
7. Is there a relationship between the **time** a person has practiced driving while having a learner's permit, and **whether or not this person passed the driving test**?
8. Can you predict a person's **favorite type of music** (classical, rock, jazz) based on his/her **IQ level**?

Role of a Variable in a Study

Learning Objectives

LO 4.19: For a data analysis situation involving two variables, identify the role of each variable in the scenario.

In most studies involving two variables, each of the variables has a role. We distinguish between:

- the **response** variable — the outcome of the study; and
- the **explanatory** variable — the variable that claims to explain, predict or affect the response.

As we mentioned earlier the variable we wish to predict is commonly called the **dependent variable**, the **outcome** variable, or the **response** variable. Any variable we are using to predict (or explain differences) in the outcome is commonly called an **explanatory variable**, an **independent variable**, a **predictor** variable, or a **covariate**.

Comment:

- Typically the **explanatory** variable is denoted by X, and the **response** variable by Y.

Now let's go back to some of the examples and classify the two relevant variables according to their roles in the study:

✓ **EXAMPLE 1:**

Is there a relationship between **gender** and **test scores** on a particular standardized test? Other ways of phrasing the same research question:

- Is performance on the test related to gender?
- Is there a gender effect on test scores?
- Are there differences in test scores between males and females?

We want to explore whether the outcome of the study — the score on a test — is affected by the test-taker's gender. Therefore:

Gender is the **explanatory** variable

Test score is the **response** variable

✓ **EXAMPLE 3:**

Is there a relationship between the **type of light** a baby sleeps with (no light, night-light, lamp) and whether or not the child develops **nearsightedness**?

In this study we explore whether the nearsightedness of a person can be explained by the type of light that person slept with as a baby. Therefore:

Light type is the **explanatory** variable

Nearsightedness is the **response** variable

✓ **EXAMPLE 5:**

How well can we predict a student's freshman year **GPA** from his/her **SAT score**?

Here we are examining whether a student's SAT score is a good predictor for the student's GPA freshman year. Therefore:

SAT score is the **explanatory** variable

GPA of freshman year is the **response** variable

✓ **EXAMPLE 7:**

Is there a relationship between the **time** a person has practiced driving while having a learner's permit, and **whether or not this person passed the driving test**?

Here we are examining whether a person's outcome on the driving test (pass/fail) can be explained by the length of time this person has practiced driving prior to the test. Therefore:

Time is the **explanatory** variable

Driving test outcome is the **response** variable

Now, using the same reasoning, the following exercise will help you to classify the two variables in the other examples.

Learn By Doing: [Role Classification](#)

✓ Many Students Wonder: Role Classification

Question: Is the role classification of variables always clear? In other words, is it always clear which of the variables is the explanatory and which is the response?

Answer: No. There are studies in which the role classification is not really clear. This mainly happens in cases when both variables are categorical or both are quantitative. An example is a study that explores the relationship between students' SAT Math and SAT Verbal scores. In cases like this, any classification choice would be fine (as long as it is consistent throughout the analysis).

Role-Type Classification

Learning Objectives

LO 4.20: Classify a data analysis situation involving two variables according to the “role-type classification.”

If we further classify each of the two relevant variables according to **type** (categorical or quantitative), we get the following 4 possibilities for “**role-type classification**”

1. Categorical explanatory and quantitative response (Case CQ)
2. Categorical explanatory and categorical response (Case CC)
3. Quantitative explanatory and quantitative response (Case QQ)
4. Quantitative explanatory and categorical response (Case QC)

This role-type classification can be summarized and easily visualized in the following table (note that the explanatory variable is always listed first):

		Response	
		Categorical	Quantitative
Explanatory	Categorical	C → C	C → Q
	Quantitative	Q → C	Q → Q

This role-type classification serves as the infrastructure for this entire section. In each of the 4 cases, different statistical tools (displays and numerical measures) should be used in order to explore the relationship between the two variables.

This suggests the following important principle:

PRINCIPLE: When confronted with a research question that involves exploring the relationship between two variables, the first and most crucial step is to determine which of the 4 cases represents the data structure of the problem. In other words, the first step should be classifying the two relevant variables according to their role and type, and only then can we determine what statistical tools should be used to analyze them.

Now let's go back to our 8 examples and determine which of the 4 cases represents the data structure of each:

✓ EXAMPLE 1:

Is there a relationship between **gender** and **test scores** on a particular standardized test? Other ways of phrasing the same research question:

- Is performance on the test related to gender?
- Is there a gender effect on test scores?
- Are there differences in test scores between males and females?

We want to explore whether the outcome of the study — the score on a test — is affected by the test-taker's gender.

Gender is the **explanatory** variable and it is **categorical**.

Test score is the **response** variable and it is **quantitative**.

Therefore this is an example of **case $C \rightarrow Q$** .

✓ **EXAMPLE 3:**

Is there a relationship between the **type of light** a baby sleeps with (no light, night-light, lamp) and whether or not the child develops **nearsightedness**?

In this study we explore whether the nearsightedness of a person can be explained by the type of light that person slept with as a baby.

Light type is the **explanatory** variable and it is **categorical**.

Nearsightedness is the **response** variable and it is **categorical**.

Therefore this is an example of **case $C \rightarrow C$** .

✓ **EXAMPLE 5:**

How well can we predict a student's freshman year **GPA** from his/her **SAT score**?

Here we are examining whether a student's SAT score is a good predictor for the student's GPA freshman year.

SAT score is the **explanatory** variable and it is **quantitative**.

GPA of freshman year is the **response** variable and it is **quantitative**.

Therefore this is an example of **case $Q \rightarrow Q$** .

✓ **EXAMPLE 7:**

Is there a relationship between the **time** a person has practiced driving while having a learner's permit, and **whether or not this person passed the driving test**?

Here we are examining whether a person's outcome on the driving test (pass/fail) can be explained by the length of time this person has practiced driving prior to the test.

Time is the **explanatory** variable and it is **quantitative**.

Driving test outcome is the **response** variable and it is **categorical**.

Therefore this is an example of **case $Q \rightarrow C$** .

Now you complete the rest...

Learn By Doing: [Role-Type Classification](#)

The remainder of this section on exploring relationships will be guided by this role-type classification. In the next three parts we will elaborate on cases $C \rightarrow Q$, $C \rightarrow C$, and $Q \rightarrow Q$. More specifically, we will learn the appropriate statistical tools (visual display and numerical measures) that will allow us to explore the relationship between the two variables in each of the cases. Case $Q \rightarrow C$ will **not** be discussed in this course, and is typically covered in more advanced courses. The section will conclude with a discussion on causal relationships.

Did I Get This?: [Role-Type Classification](#)

[Role-Type Classification](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by LibreTexts.