INDIAN INSTITUTE OF TECHNOLOGY MADRAS

सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Statistics for Data Science -1

Lecture 4.2: Association between two categorical variables-Introduction

Usha Mohan

Indian Institute of Technology Madras

Learning objectives

Learning objectives

1. Use of two-way contingency tables to understand association between two categorical variables.

Learning objectives

1. Use of two-way contingency tables to understand association between two categorical variables.
2. Understand association between numerical variables through scatter plots; compute and interpret correlation.

Learning objectives

1. Use of two-way contingency tables to understand association between two categorical variables.
2. Understand association between numerical variables through scatter plots; compute and interpret correlation.
3. Understand relationship between a categorical and numerical variable.

Introduction

- To understand the association between two categorical variables.
- Learn how to construct two-way contingency table.
- Learn concept of relative row/column frequencies and how to use them to determine whether there is an association between the categorical variables.

Example 1: Gender versus use of smartphone

Example 1: Gender versus use of smartphone

- A market research firm is interested in finding out whether ownership of a smartphone is associated with gender of a student. In other words, they want to find out whether more females own a smartphone while compared to males, or whether owning a smartphone is independent of gender.

Example 1: Gender versus use of smartphone

- A market research firm is interested in finding out whether ownership of a smartphone is associated with gender of a student. In other words, they want to find out whether more females own a smartphone while compared to males, or whether owning a smartphone is independent of gender.
- To answer this question, a group of 100 college going children were surveyed about whether they owned a smart phone or not.

Example 1: Gender versus use of smartphone

- A market research firm is interested in finding out whether ownership of a smartphone is associated with gender of a student. In other words, they want to find out whether more females own a smartphone while compared to males, or whether owning a smartphone is independent of gender.
- To answer this question, a group of 100 college going children were surveyed about whether they owned a smart phone or not.
- The categorical variables in this example are

Example 1: Gender versus use of smartphone

- A market research firm is interested in finding out whether ownership of a smartphone is associated with gender of a student. In other words, they want to find out whether more females own a smartphone while compared to males, or whether owning a smartphone is independent of gender.
- To answer this question, a group of 100 college going children were surveyed about whether they owned a smart phone or not.
- The categorical variables in this example are
  - Gender: Male, Female (2 categories)- Nominal variable

Example 1: Gender versus use of smartphone

- A market research firm is interested in finding out whether ownership of a smartphone is associated with gender of a student. In other words, they want to find out whether more females own a smartphone while compared to males, or whether owning a smartphone is independent of gender.
- To answer this question, a group of 100 college going children were surveyed about whether they owned a smart phone or not.
- The categorical variables in this example are
    - Gender: Male, Female (2 categories)- Nominal variable
    - Own a smartphone: Yes, No (2 categories)- Nominal variable

Example 1: Gender versus use of smartphone-summarize data

Example 1: Gender versus use of smartphone-summarize data

- We have the following summary statistics

Example 1: Gender versus use of smartphone-summarize data

- We have the following summary statistics
  1. There are 44 female and 56 male students

Example 1: Gender versus use of smartphone-summarize data

- We have the following summary statistics
  1. There are 44 female and 56 male students
  2. 76 students owned a smartphone, 24 did not own.

Example 1: Gender versus use of smartphone-summarize data

- We have the following summary statistics
  1. There are 44 female and 56 male students
  2. 76 students owned a smartphone, 24 did not own.
  3. 34 female students owned a smartphone, 42 male students owned a smartphone.

Example 1: Gender versus use of smartphone-summarize data

- We have the following summary statistics
  1. There are 44 female and 56 male students
  2. 76 students owned a smartphone, 24 did not own.
  3. 34 female students owned a smartphone, 42 male students owned a smartphone.
- The data given in the example can be organized using a two-way table, referred to as a contingency table.

Example 1: Gender versus use of smartphone-summarize data

- We have the following summary statistics
  1. There are 44 female and 56 male students
  2. 76 students owned a smartphone, 24 did not own.
  3. 34 female students owned a smartphone, 42 male students owned a smartphone.
- The data given in the example can be organized using a two-way table, referred to as a contingency table.

	Own a smartphone		
Gender	**No**	**Yes**	**Row total**
Female	10	34	**44**
Male	14	42	**56**
Column total	**24**	**76**	**100**

Contingency table using google sheets

Step 1 Choose the columns of the variables for which you seek an association.

Step 2 Go to Data-click on Pivot table option

Step 3 Click on create option in the pivot table- it will open the pivot table editor:

- 3.1 Under the Rows tab, click on the first categorical variable.
- 3.2 Under the columns tab, click on the second categorical variable.
- 3.3 Under the values tab, click on either of the variables and then click on the COUNTA tab under "summarize by" tab.

Example 2: Income versus use of smartphone

Example 2: Income versus use of smartphone

- A market research firm is interested in finding out whether ownership of a smartphone is associated with income of an individual. In other words, they want to find out whether income is associated with ownership of a smartphone.

Example 2: Income versus use of smartphone

- A market research firm is interested in finding out whether ownership of a smartphone is associated with income of an individual. In other words, they want to find out whether income is associated with ownership of a smartphone.
- To answer this question, a group of 100 randomly picked individuals were surveyed about whether they owned a smart phone or not.

Example 2: Income versus use of smartphone

- A market research firm is interested in finding out whether ownership of a smartphone is associated with income of an individual. In other words, they want to find out whether income is associated with ownership of a smartphone.
- To answer this question, a group of 100 randomly picked individuals were surveyed about whether they owned a smart phone or not.
- The categorical variables in this example are

Example 2: Income versus use of smartphone

- A market research firm is interested in finding out whether ownership of a smartphone is associated with income of an individual. In other words, they want to find out whether income is associated with ownership of a smartphone.
- To answer this question, a group of 100 randomly picked individuals were surveyed about whether they owned a smart phone or not.
- The categorical variables in this example are
  - Income: Low, Medium, High (3 categories) -Ordinal variable

Example 2: Income versus use of smartphone

- A market research firm is interested in finding out whether ownership of a smartphone is associated with income of an individual. In other words, they want to find out whether income is associated with ownership of a smartphone.
- To answer this question, a group of 100 randomly picked individuals were surveyed about whether they owned a smart phone or not.
- The categorical variables in this example are
  - Income: Low, Medium, High (3 categories) -Ordinal variable
  - Own a smartphone: Yes, No (2 categories) - Nominal variable

Example 2: Contingency table

Example 2: Contingency table

- We have the following summary statistics
  1. There are 20 High income, 66 medium income, and 14 low income participants.
  2. 62 participants owned a smartphone, 38 did not own.
  3. 18 High income participants owned a smartphone, 39 Medium income participants owned a smartphone, and 5 Low income participants owned a smartphone.

Example 2: Contingency table

- We have the following summary statistics
  1. There are 20 High income, 66 medium income, and 14 low income participants.
  2. 62 participants owned a smartphone, 38 did not own.
  3. 18 High income participants owned a smartphone, 39 Medium income participants owned a smartphone, and 5 Low income participants owned a smartphone.
- The contingency table corresponding to the data is given below.

	Own a smartphone		
Income level	**No**	**Yes**	**Row total**
High	2	18	**20**
Medium	27	39	**66**
Low	9	5	**14**
Column total	**38**	**62**	**100**

Section summary

- Organize bivariate categorical data into a two-way table-contingency table.
- If data is ordinal, maintain order of the variable in the table