

## Finding Relationships Among Variables

BUS 230: Business and Economic Research and  
Communication

- Specific goals:
  - Re-familiarize ourselves with basic statistics ideas: sampling distributions, hypothesis tests, p-values.
  - Be able to distinguish different types of data and prescribe appropriate statistical methods.
  - Conduct a number of hypothesis tests using methods appropriate for questions involving only one or two variables.
- Learning objectives:
  - LO2: Interpret data using statistical analysis.
  - LO2.3: Formulate conclusions and recommendations based upon statistical results.

# Goals

- Specific goals:
  - Re-familiarize ourselves with basic statistics ideas: sampling distributions, hypothesis tests, p-values.
  - Be able to distinguish different types of data and prescribe appropriate statistical methods.
  - Conduct a number of hypothesis tests using methods appropriate for questions involving only one or two variables.
- Learning objectives:
  - LO2: Interpret data using statistical analysis.
  - LO2.3: Formulate conclusions and recommendations based upon statistical results.

# What to Look For

- For each test, remember the following:
  - In plain English, be able to describe the purpose of the test.
  - Know whether the test is a parametric test or a non-parametric test.
  - Know the null and alternative hypotheses.
- When choosing a test to answer a research question, ask yourself:
  - ① What is your research question?
  - ② How many variables do you have?
  - ③ Are you looking for differences or other relationship?
  - ④ If you are looking for differences, are your observations independent or paired?

# What to Look For

- For each test, remember the following:
  - In plain English, be able to describe the purpose of the test.
  - Know whether the test is a parametric test or a non-parametric test.
  - Know the null and alternative hypotheses.
- When choosing a test to answer a research question, ask yourself:
  - ① What is your research question?
  - ② How many variables do you have?
  - ③ Are you looking for differences or other relationship?
  - ④ If you are looking for differences, are your observations independent or paired?

# What to Look For

- For each test, remember the following:
  - In plain English, be able to describe the purpose of the test.
  - Know whether the test is a parametric test or a non-parametric test.
  - Know the null and alternative hypotheses.
- When choosing a test to answer a research question, ask yourself:
  - What is your research question?
  - How many variables do you have?
  - Are you looking for differences or other relationship?
  - If you are looking for differences, are your observations independent or paired?

# What to Look For

- For each test, remember the following:
  - In plain English, be able to describe the purpose of the test.
  - Know whether the test is a parametric test or a non-parametric test.
  - Know the null and alternative hypotheses.
- When choosing a test to answer a research question, ask yourself:
  - What is your research question?
  - How many variables do you have?
  - Are you looking for differences or other relationship?
  - If you are looking for differences, are your observations independent or paired?

# What to Look For

- For each test, remember the following:
  - In plain English, be able to describe the purpose of the test.
  - Know whether the test is a parametric test or a non-parametric test.
  - Know the null and alternative hypotheses.
- When choosing a test to answer a research question, ask yourself:
  - 1 What is your research question?
  - 2 How many variables do you have?
  - 3 Are you looking for differences or other relationship?
  - 4 If you are looking for differences, are your observations independent or paired?



# What to Look For

- For each test, remember the following:
  - In plain English, be able to describe the purpose of the test.
  - Know whether the test is a parametric test or a non-parametric test.
  - Know the null and alternative hypotheses.
- When choosing a test to answer a research question, ask yourself:
  - 1 What is your research question?
  - 2 How many variables do you have?
  - 3 Are you looking for differences or other relationship?
  - 4 If you are looking for differences, are your observations independent or paired?

# What to Look For

- For each test, remember the following:
  - In plain English, be able to describe the purpose of the test.
  - Know whether the test is a parametric test or a non-parametric test.
  - Know the null and alternative hypotheses.
- When choosing a test to answer a research question, ask yourself:
  - 1 What is your research question?
  - 2 How many variables do you have?
  - 3 Are you looking for differences or other relationship?
  - 4 If you are looking for differences, are your observations independent or paired?

# What to Look For

- For each test, remember the following:
  - In plain English, be able to describe the purpose of the test.
  - Know whether the test is a parametric test or a non-parametric test.
  - Know the null and alternative hypotheses.
- When choosing a test to answer a research question, ask yourself:
  - 1 What is your research question?
  - 2 How many variables do you have?
  - 3 Are you looking for differences or other relationship?
  - 4 If you are looking for differences, are your observations independent or paired?

# What to Look For

- For each test, remember the following:
  - In plain English, be able to describe the purpose of the test.
  - Know whether the test is a parametric test or a non-parametric test.
  - Know the null and alternative hypotheses.
- When choosing a test to answer a research question, ask yourself:
  - 1 What is your research question?
  - 2 How many variables do you have?
  - 3 Are you looking for differences or other relationship?
  - 4 If you are looking for differences, are your observations independent or paired?

# Correlation

- A **correlation** exists between two variables when one of them is related to the other in some way, such that there is **co-movement**.
- The **Pearson linear correlation coefficient** is a measure of the strength of the linear relationship between two variables.
  - Parametric test!
  - Null hypothesis: there is zero linear correlation between two variables.
  - Alternative hypothesis: there is a linear correlation (either positive or negative) between two variables.
- Spearman's Rank Test
  - Non-parametric test.
  - Behind the scenes - replaces actual data with their *rank*, computes the Pearson using ranks.
  - Same hypotheses.

# Correlation

- A **correlation** exists between two variables when one of them is related to the other in some way, such that there is **co-movement**.
- The **Pearson linear correlation coefficient** is a measure of the strength of the linear relationship between two variables.
  - Parametric test!
  - Null hypothesis: there is zero linear correlation between two variables.
  - Alternative hypothesis: there is a linear correlation (either positive or negative) between two variables.
- Spearman's Rank Test
  - Non-parametric test.
  - Behind the scenes - replaces actual data with their *rank*, computes the Pearson using ranks.
  - Same hypotheses.

# Correlation

- A **correlation** exists between two variables when one of them is related to the other in some way, such that there is **co-movement**.
- The **Pearson linear correlation coefficient** is a measure of the strength of the linear relationship between two variables.
  - Parametric test!
  - Null hypothesis: there is zero linear correlation between two variables.
  - Alternative hypothesis: there is a linear correlation (either positive or negative) between two variables.
- Spearman's Rank Test
  - Non-parametric test.
  - Behind the scenes - replaces actual data with their *rank*, computes the Pearson using ranks.
  - Same hypotheses.

# Correlation

- A **correlation** exists between two variables when one of them is related to the other in some way, such that there is **co-movement**.
- The **Pearson linear correlation coefficient** is a measure of the strength of the linear relationship between two variables.
  - Parametric test!
  - Null hypothesis: there is zero linear correlation between two variables.
  - Alternative hypothesis: there is a linear correlation (either positive or negative) between two variables.
- Spearman's Rank Test
  - Non-parametric test.
  - Behind the scenes - replaces actual data with their *rank*, computes the Pearson using ranks.
  - Same hypotheses.



# Correlation

- A **correlation** exists between two variables when one of them is related to the other in some way, such that there is **co-movement**.
- The **Pearson linear correlation coefficient** is a measure of the strength of the linear relationship between two variables.
  - Parametric test!
  - Null hypothesis: there is zero linear correlation between two variables.
  - Alternative hypothesis: there is a linear correlation (either positive or negative) between two variables.
- Spearman's Rank Test
  - Non-parametric test.
  - Behind the scenes - replaces actual data with their *rank*, computes the Pearson using ranks.
  - Same hypotheses.

# Correlation

- A **correlation** exists between two variables when one of them is related to the other in some way, such that there is **co-movement**.
- The **Pearson linear correlation coefficient** is a measure of the strength of the linear relationship between two variables.
  - Parametric test!
  - Null hypothesis: there is zero linear correlation between two variables.
  - Alternative hypothesis: there is a linear correlation (either positive or negative) between two variables.
- Spearman's Rank Test
  - Non-parametric test.
  - Behind the scenes - replaces actual data with their *rank*, computes the Pearson using ranks.
  - Same hypotheses.

# Correlation

- A **correlation** exists between two variables when one of them is related to the other in some way, such that there is **co-movement**.
- The **Pearson linear correlation coefficient** is a measure of the strength of the linear relationship between two variables.
  - Parametric test!
  - Null hypothesis: there is zero linear correlation between two variables.
  - Alternative hypothesis: there is a linear correlation (either positive or negative) between two variables.
- Spearman's Rank Test
  - Non-parametric test.
  - Behind the scenes - replaces actual data with their *rank*, computes the Pearson using ranks.
  - Same hypotheses.

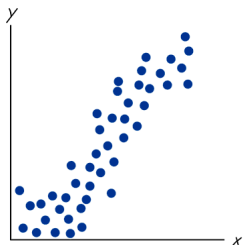
# Correlation

- A **correlation** exists between two variables when one of them is related to the other in some way, such that there is **co-movement**.
- The **Pearson linear correlation coefficient** is a measure of the strength of the linear relationship between two variables.
  - Parametric test!
  - Null hypothesis: there is zero linear correlation between two variables.
  - Alternative hypothesis: there is a linear correlation (either positive or negative) between two variables.
- Spearman's Rank Test
  - Non-parametric test.
  - Behind the scenes - replaces actual data with their *rank*, computes the Pearson using ranks.
  - Same hypotheses.

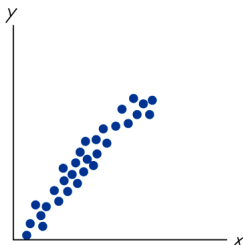
# Correlation

- A **correlation** exists between two variables when one of them is related to the other in some way, such that there is **co-movement**.
- The **Pearson linear correlation coefficient** is a measure of the strength of the linear relationship between two variables.
  - Parametric test!
  - Null hypothesis: there is zero linear correlation between two variables.
  - Alternative hypothesis: there is a linear correlation (either positive or negative) between two variables.
- Spearman's Rank Test
  - Non-parametric test.
  - Behind the scenes - replaces actual data with their *rank*, computes the Pearson using ranks.
  - Same hypotheses.

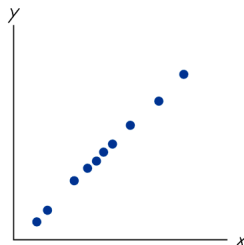
# Positive linear correlation



(a) Positive correlation between  $x$  and  $y$



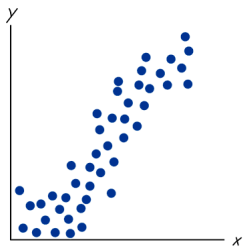
(b) Strong positive correlation between  $x$  and  $y$



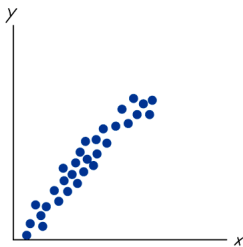
(c) Perfect positive correlation between  $x$  and  $y$

- Positive correlation: two variables move in the same direction.
- Stronger the correlation: closer the correlation coefficient is to 1.
- Perfect positive correlation:  $\rho = 1$

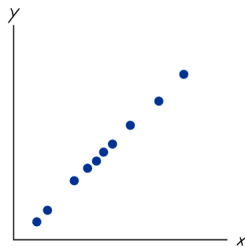
# Positive linear correlation



(a) Positive correlation between  $x$  and  $y$



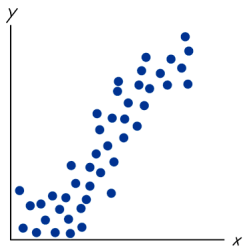
(b) Strong positive correlation between  $x$  and  $y$



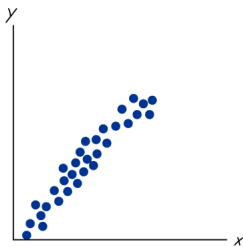
(c) Perfect positive correlation between  $x$  and  $y$

- Positive correlation: two variables move in the same direction.
- Stronger the correlation: closer the correlation coefficient is to 1.
- Perfect positive correlation:  $\rho = 1$

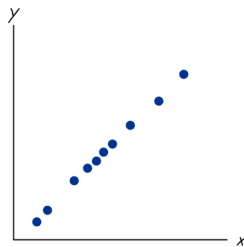
# Positive linear correlation



(a) Positive correlation between  $x$  and  $y$



(b) Strong positive correlation between  $x$  and  $y$

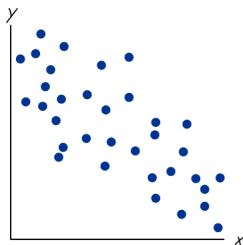


(c) Perfect positive correlation between  $x$  and  $y$

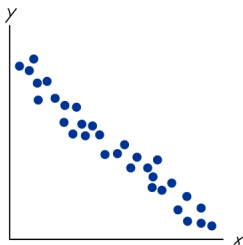
- Positive correlation: two variables move in the same direction.
- Stronger the correlation: closer the correlation coefficient is to 1.
- Perfect positive correlation:  $\rho = 1$



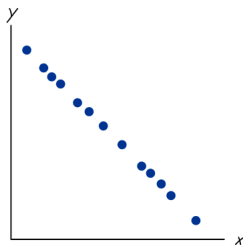
# Negative linear correlation



(d) Negative correlation between  $x$  and  $y$



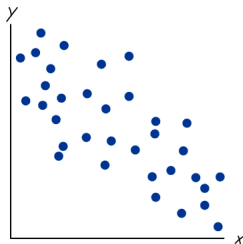
(e) Strong negative correlation between  $x$  and  $y$



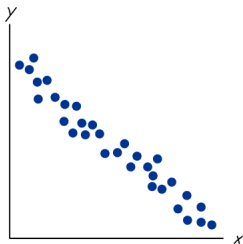
(f) Perfect negative correlation between  $x$  and  $y$

- Negative correlation: two variables move in opposite directions.
- Stronger the correlation: closer the correlation coefficient is to  $-1$ .
- Perfect negative correlation:  $\rho = -1$

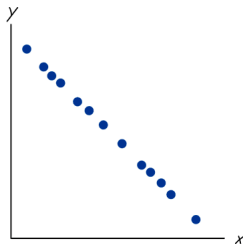
# Negative linear correlation



(d) Negative correlation between  $x$  and  $y$



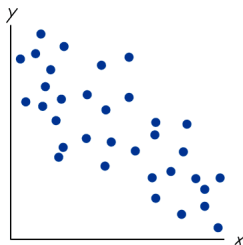
(e) Strong negative correlation between  $x$  and  $y$



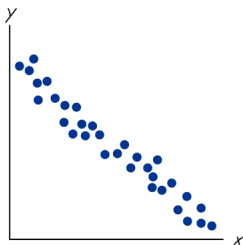
(f) Perfect negative correlation between  $x$  and  $y$

- Negative correlation: two variables move in opposite directions.
- Stronger the correlation: closer the correlation coefficient is to  $-1$ .
- Perfect negative correlation:  $\rho = -1$

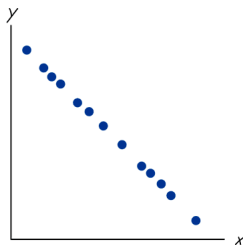
# Negative linear correlation



**(d)** Negative correlation between  $x$  and  $y$



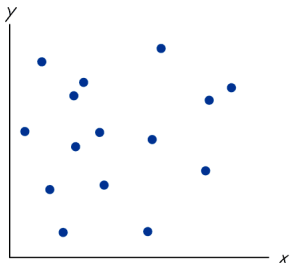
**(e)** Strong negative correlation between  $x$  and  $y$



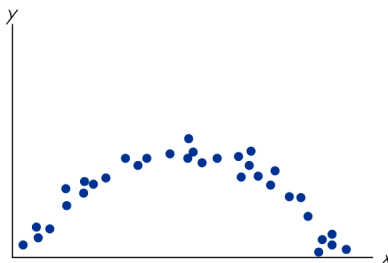
**(f)** Perfect negative correlation between  $x$  and  $y$

- Negative correlation: two variables move in opposite directions.
- Stronger the correlation: closer the correlation coefficient is to  $-1$ .
- Perfect negative correlation:  $\rho = -1$

## No linear correlation



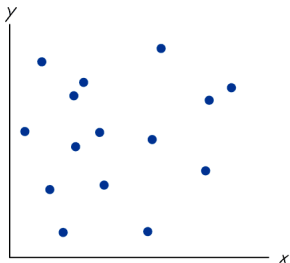
(g) No correlation between  $x$  and  $y$



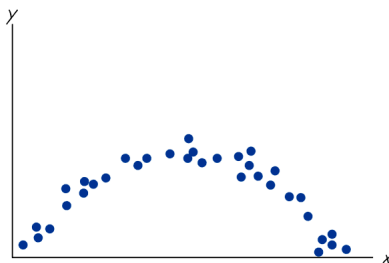
(h) Nonlinear relationship between  $x$  and  $y$

- Panel (g): no relationship at all.
- Panel (h): strong relationship, but not a *linear* relationship.
  - Cannot use regular correlation to detect this.

# No linear correlation



(g) No correlation between  $x$  and  $y$



(h) Nonlinear relationship between  $x$  and  $y$

- Panel (g): no relationship at all.
- Panel (h): strong relationship, but not a *linear* relationship.
  - Cannot use regular correlation to detect this.

# Chi-Squared Test for Independence

- Used to determine if two categorical variables (eg: nominal) are related.
- Example: Suppose a hotel manager surveys guest who indicate they will not return:

Reason for Stay	Reason for Not Returning		
	Price	Location	Amenities
Personal/Vacation	56	49	0
Business	20	47	27

- Data in the table are always frequencies that fall into individual categories.
- Could use this table to test if two variables are independent.

# Chi-Squared Test for Independence

- Used to determine if two categorical variables (eg: nominal) are related.
- Example: Suppose a hotel manager surveys guest who indicate they will not return:

Reason for Stay	Reason for Not Returning		
	Price	Location	Amenities
Personal/Vacation	56	49	0
Business	20	47	27

- Data in the table are always frequencies that fall into individual categories.
- Could use this table to test if two variables are independent.

# Chi-Squared Test for Independence

- Used to determine if two categorical variables (eg: nominal) are related.
- Example: Suppose a hotel manager surveys guest who indicate they will not return:

Reason for Not Returning

Reason for Stay	Price	Location	Amenities
Personal/Vacation	56	49	0
Business	20	47	27

- Data in the table are always frequencies that fall into individual categories.
- Could use this table to test if two variables are independent.



# Chi-Squared Test for Independence

- Used to determine if two categorical variables (eg: nominal) are related.
- Example: Suppose a hotel manager surveys guest who indicate they will not return:

Reason for Not Returning

Reason for Stay	Price	Location	Amenities
Personal/Vacation	56	49	0
Business	20	47	27

- Data in the table are always frequencies that fall into individual categories.
- Could use this table to test if two variables are independent.

# Chi-Squared Test for Independence

- Used to determine if two categorical variables (eg: nominal) are related.
- Example: Suppose a hotel manager surveys guest who indicate they will not return:

Reason for Not Returning

Reason for Stay	Price	Location	Amenities
Personal/Vacation	56	49	0
Business	20	47	27

- Data in the table are always frequencies that fall into individual categories.
- Could use this table to test if two variables are independent.

# Test of independence

- **Null hypothesis:** there is no relationship between the row variable and the column variable (independent)
- **Alternative hypothesis:** There is a relationship between the row variable and the column variable (dependent).
- Test statistic:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

- $O$ : observed frequency in a cell from the contingency table.
- $E$ : expected frequency computed with the *assumption that the variables are independent*.
- Large  $\chi^2$  values indicate variables are dependent (reject the null hypothesis).

# Test of independence

- **Null hypothesis:** there is no relationship between the row variable and the column variable (independent)
- **Alternative hypothesis:** There is a relationship between the row variable and the column variable (dependent).
- Test statistic:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

- $O$ : observed frequency in a cell from the contingency table.
- $E$ : expected frequency computed with the *assumption that the variables are independent*.
- Large  $\chi^2$  values indicate variables are dependent (reject the null hypothesis).

# Test of independence

- **Null hypothesis:** there is no relationship between the row variable and the column variable (independent)
- **Alternative hypothesis:** There is a relationship between the row variable and the column variable (dependent).
- Test statistic:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

- $O$ : observed frequency in a cell from the contingency table.
- $E$ : expected frequency computed with the *assumption that the variables are independent*.
- Large  $\chi^2$  values indicate variables are dependent (reject the null hypothesis).

# Test of independence

- **Null hypothesis:** there is no relationship between the row variable and the column variable (independent)
- **Alternative hypothesis:** There is a relationship between the row variable and the column variable (dependent).
- Test statistic:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

- $O$ : observed frequency in a cell from the contingency table.
- $E$ : expected frequency computed with the *assumption that the variables are independent*.
- Large  $\chi^2$  values indicate variables are dependent (reject the null hypothesis).

# Test of independence

- **Null hypothesis:** there is no relationship between the row variable and the column variable (independent)
- **Alternative hypothesis:** There is a relationship between the row variable and the column variable (dependent).
- Test statistic:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

- $O$ : observed frequency in a cell from the contingency table.
- $E$ : expected frequency computed with the *assumption that the variables are independent*.
- Large  $\chi^2$  values indicate variables are dependent (reject the null hypothesis).

# Test of independence

- **Null hypothesis:** there is no relationship between the row variable and the column variable (independent)
- **Alternative hypothesis:** There is a relationship between the row variable and the column variable (dependent).
- Test statistic:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

- $O$ : observed frequency in a cell from the contingency table.
- $E$ : expected frequency computed with the *assumption that the variables are independent*.
- Large  $\chi^2$  values indicate variables are dependent (reject the null hypothesis).



# Test of independence

- **Null hypothesis:** there is no relationship between the row variable and the column variable (independent)
- **Alternative hypothesis:** There is a relationship between the row variable and the column variable (dependent).
- Test statistic:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

- $O$ : observed frequency in a cell from the contingency table.
- $E$ : expected frequency computed with the *assumption that the variables are independent*.
- Large  $\chi^2$  values indicate variables are dependent (reject the null hypothesis).