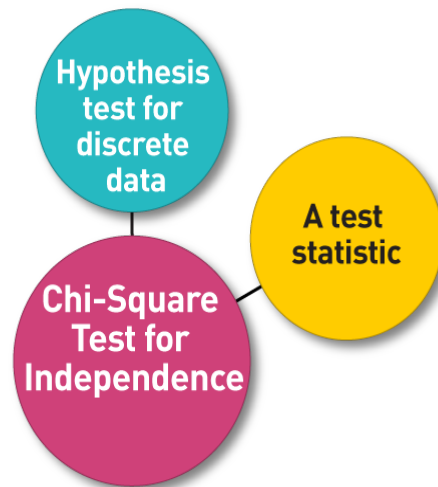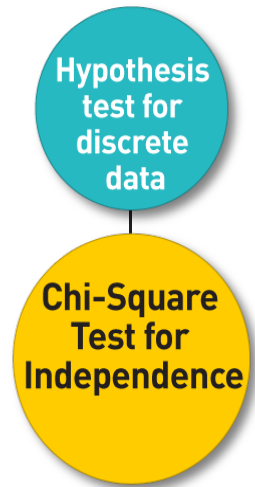# 4.3 Chi-Square Test of Independence: What Is It?

MBC 638

Data Analysis and Decision Making
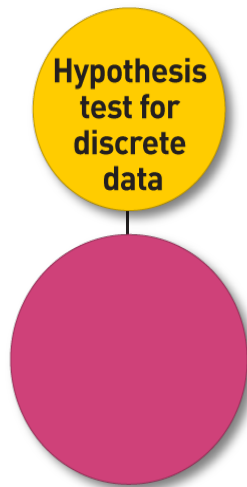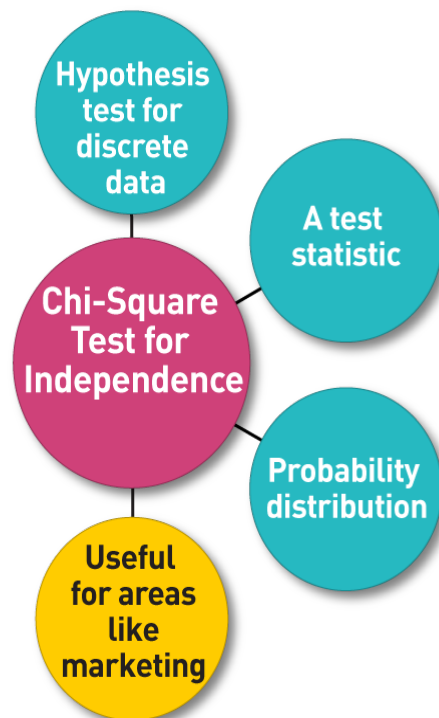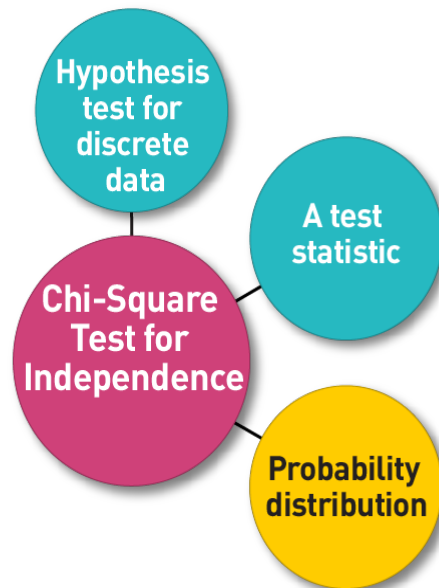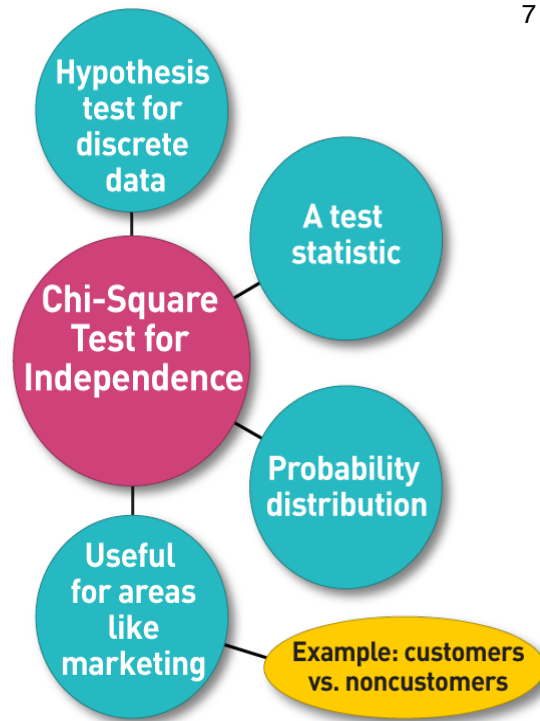
Hypothesis test for discrete data

Hypothesis test for discrete data

Chi-Square Test for Independence

Hypothesis test for discrete data

Chi-Square Test for Independence

A test statistic

Hypothesis test for discrete data

A test statistic

Chi-Square Test for Independence

Probability distribution

Useful for areas like marketing

Example: customers vs. noncustomers

Chi-Square Test for Independence

- χ^2 **test for independence:** a procedure used to determine if two variables are related or are statistically independent

Chi-Square Test for Independence

- **χ^2 test for independence:** a procedure used to determine if two variables are related or are statistically independent

- Convention:
  - H_0:

  - H_a:

Chi-Square Test for Independence

- **χ^2 test for independence:** a procedure used to determine if two variables are related or are statistically independent

- Convention:
  - H_0:

  - H_a: Categorical Variable 1 and Categorical Variable 2 are not independent (i.e., there *is* a relationship).
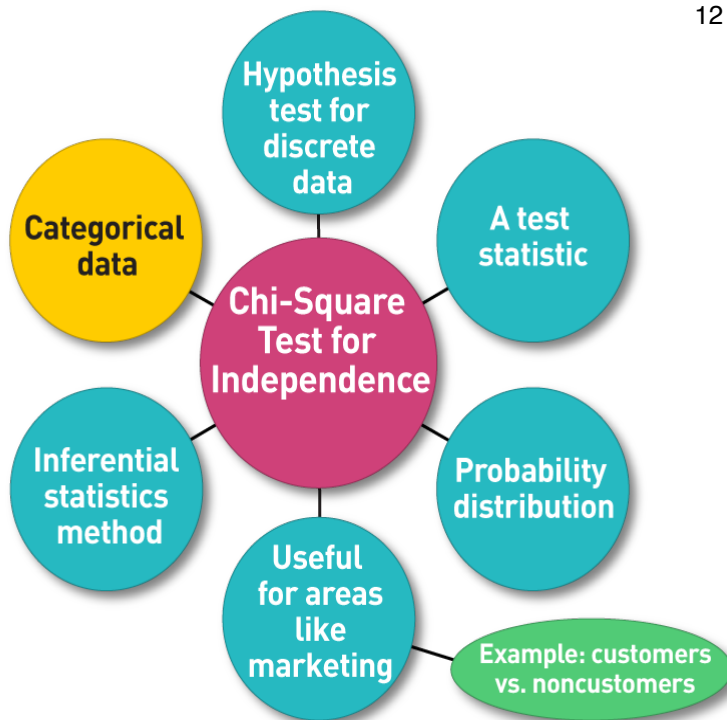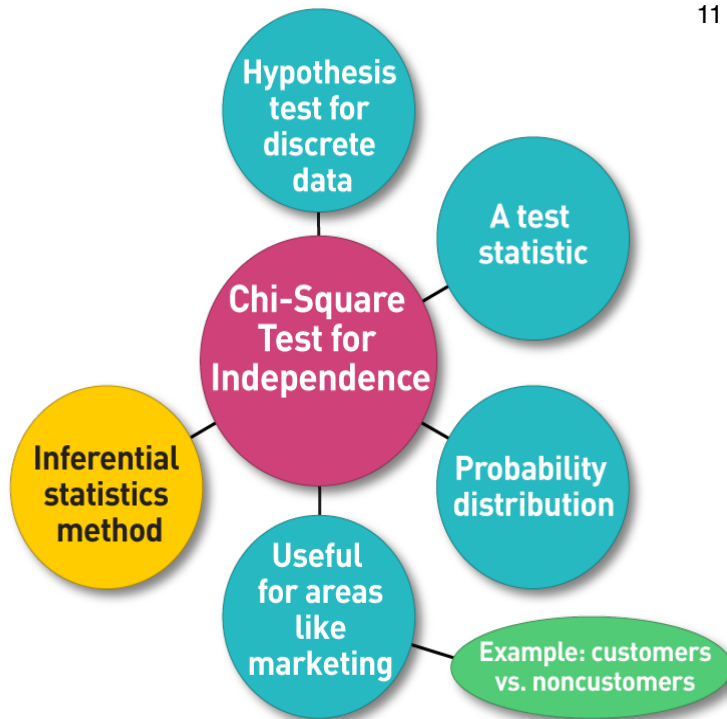
Chi-Square Test for Independence

- **χ^2 test for independence:** a procedure used to determine if two variables are related or are statistically independent

- Convention:
  - H_0: Categorical Variable 1 and Categorical Variable 2 are independent (i.e., there is *no* relationship).

  - H_a: Categorical Variable 1 and Categorical Variable 2 are not independent (i.e., there *is* a relationship).

Chi-Square Test for Independence

- **χ^2 test for independence:** a procedure used to determine if two variables are related or are statistically independent
- Convention:
  - H_0: Categorical Variable 1 and Categorical Variable 2 are independent (i.e., there is *no* relationship).
  - H_a: Categorical Variable 1 and Categorical Variable 2 are not independent (i.e., there *is* a relationship).
- How it works:

Chi-Square Test for Independence

- **χ^2 test for independence:** a procedure used to determine if two variables are related or are statistically independent
- Convention:
  - H_0: Categorical Variable 1 and Categorical Variable 2 are independent (i.e., there is *no* relationship).
  - H_a: Categorical Variable 1 and Categorical Variable 2 are not independent (i.e., there *is* a relationship).
- How it works:
  - Compares "observed" counts and "expected" counts or frequencies

Chi-Square Test for Independence

- χ^2 **test for independence:** a procedure used to determine if two variables are related or are statistically independent
- Convention:
    - H_0: Categorical Variable 1 and Categorical Variable 2 are independent (i.e., there is *no* relationship).
    - H_a: Categorical Variable 1 and Categorical Variable 2 are not independent (i.e., there *is* a relationship).
- How it works:
    - Compares "observed" counts and "expected" counts or frequencies
    - Does not give kind (positive/negative) or intensity of relationship

Setup for Analysis of Two-Way Tables

H_0: Categorical Variable 1 and Categorical Variable 2 are independent.
H_a: Categorical Variable 1 and Categorical Variable 2 are not independent.

Variable 2

Totals

Variable 1

Row total

Row total

Totals   Col. total   Col. total   Col. total   Col. total   Col. total   Col. total   Col. total   Grand total

Setup for Analysis of Two-Way Tables

H_0: Categorical Variable 1 and Categorical Variable 2 are independent.
H_a: Categorical Variable 1 and Categorical Variable 2 are not independent.

Variable 2

Totals

Variable 1

Row total

Row total

Totals   Col. total   Col. total   Col. total   Col. total   Col. total   Col. total   Col. total   Grand total

- Put your data in a two-way table.

Setup for Analysis of Two-Way Tables

H_0: Categorical Variable 1 and Categorical Variable 2 are independent.
H_a: Categorical Variable 1 and Categorical Variable 2 are not independent.

Variable 2

Totals

Variable 1

| | | | | | | | Row total |
| | | | | | | | Row total |

Totals

| Col. total | Col. total | Col. total | Col. total | Col. total | Col. total | Col. total | Grand total |

- Put your data in a two-way table.

- Depending on the variables, use two or more columns or rows.

Setup for Analysis of Two-Way Tables

H_0: Categorical Variable 1 and Categorical Variable 2 are independent.
H_a: Categorical Variable 1 and Categorical Variable 2 are not independent.

Example: Does day of week affect car sales?

Variable 2

Totals

Variable 1

| | | | | | | | Row total |
| | | | | | | | Row total |

Totals

| Col. total | Col. total | Col. total | Col. total | Col. total | Col. total | Col. total | Grand total |

- Put your data in a two-way table.

- Depending on the variables, use two or more columns or rows.

Setup for Analysis of Two-Way Tables

H_0: Categorical Variable 1 and Categorical Variable 2 are independent.
H_a: Categorical Variable 1 and Categorical Variable 2 are not independent.

Example: Does day of week affect car sales?

Variable 2

Totals

Variable 1
Customer lead resulting in:

Sale of car								Row total
Not selling a car								Row total
Totals	Col. total	Col. total	Col. total	Col. total	Col. total	Col. total	Col. total	Grand total

- Put your data in a two-way table.

- Depending on the variables, use two or more columns or rows.

Setup for Analysis of Two-Way Tables

H_0: Categorical Variable 1 and Categorical Variable 2 are independent.
H_a: Categorical Variable 1 and Categorical Variable 2 are not independent.

Example: Does day of week affect car sales?

Variable 2

Variable 1
Customer lead resulting in:

	MON	TUES	WED	THUR	FRI	SAT	SUN	Totals
Sale of car								Row total
Not selling a car								Row total
Totals	Col. total	Col. total	Col. total	Col. total	Col. total	Col. total	Col. total	Grand total

- Put your data in a two-way table.

- Depending on the variables, use two or more columns or rows.

Setup for Analysis of Two-Way Tables

H_0: Categorical Variable 1 and Categorical Variable 2 are independent.
H_a: Categorical Variable 1 and Categorical Variable 2 are not independent.

Example: Does day of week affect car sales?

		Variable 2							**Totals**
		MON	TUES	WED	THUR	FRI	SAT	SUN	
Variable 1 Customer lead resulting in:	Sale of car	Count data	Count data	Count data	Count data	Count data	Count data	Count data	Row total
	Not selling a car	Count data	Count data	Count data	Count data	Count data	Count data	Count data	Row total
	Totals	Col. total	Col. total	Col. total	Col. total	Col. total	Col. total	Col. total	Grand total

- Put your data in a two-way table.
- Depending on the variables, use two or more columns or rows.

Setup for Analysis of Two-Way Tables

H_0: Categorical Variable 1 and Categorical Variable 2 are independent.
H_a: Categorical Variable 1 and Categorical Variable 2 are not independent.

Example: Does day of week affect car sales?

		Variable 2							**Totals**
		MON	TUES	WED	THUR	FRI	SAT	SUN	
Variable 1 Customer lead resulting in:	Sale of car	Count data	Count data	Count data	Count data	Count data	Count data	Count data	Row total
	Not selling a car	Count data	Count data	Count data	Count data	Count data	Count data	Count data	Row total
	Totals	Col. total	Col. total	Col. total	Col. total	Col. total	Col. total	Col. total	Grand total

- Put your data in a two-way table.
- Depending on the variables, use two or more columns or rows.
- In each cell, count the frequency of simultaneous occurrence.

Setup for Analysis of Two-Way Tables

H_0: Categorical Variable 1 and Categorical Variable 2 are independent.
H_a: Categorical Variable 1 and Categorical Variable 2 are not independent.

Example: Does day of week affect car sales?

		Variable 2							**Totals**
		MON	TUES	WED	THUR	FRI	SAT	SUN	
Variable 1 Customer lead resulting in:	Sale of car	Count data	Count data	Count data	Count data	Count data	Count data	Count data	Row total
	Not selling a car	Count data	Count data	Count data	Count data	Count data	Count data	Count data	Row total
Totals		Col. total	Col. total	Col. total	Col. total	Col. total	Col. total	Col. total	Grand total

- Put your data in a two-way table.

- Depending on the variables, use two or more columns or rows.

- In each cell, count the frequency of simultaneous occurrence.
  - Discrete data: need a lot of data (> 5 counts per cell)

Setup for Analysis of Two-Way Tables

H_0: Categorical Variable 1 and Categorical Variable 2 are independent.
H_a: Categorical Variable 1 and Categorical Variable 2 are not independent.

Example: Does day of week affect car sales?

Variable 2

		MON	TUES	WED	THUR	FRI	SAT	SUN	**Totals**
Variable 1 Customer lead resulting in:	Sale of car	Count data	Count data	Count data	Count data	Count data	Count data	Count data	Row total
	Not selling a car	Count data	Count data	Count data	Count data	Count data	Count data	Count data	Row total
	Totals	Col. total	Col. total	Col. total	Col. total	Col. total	Col. total	Col. total	Grand total

- Put your data in a two-way table.

- Depending on the variables, use two or more columns or rows.

- In each cell, count the frequency of simultaneous occurrence.

  - Discrete data: need a lot of data (> 5 counts per cell)

  - No exact sample size specified