Advanced Data Visualization Lecture Notes (Lecture 8 - 2 Hours)

Interaction Effects in data visualization

Duration: 2 hours

---

1. Introduction to Interaction Effects (10 mins)

Definition: Interaction effects occur when the effect of one independent variable on the dependent variable changes depending on the level of another independent variable. It means that the variables "interact" in determining the outcome, rather than acting independently.

- **Example:** In a study analyzing the effect of exercise and diet on weight loss, the impact of exercise may differ depending on the type of diet, indicating an interaction effect.

Why Interaction Effects Matter:

- Helps in understanding complex relationships between variables.
- Provides a deeper insight into how multiple factors affect an outcome together.
- Often used in fields like psychology, economics, and social sciences.

---

2. Types of Interaction Effects (10 mins)

a. Two-Way Interaction: Occurs when two independent variables interact to affect the dependent variable.

b. Three-Way Interaction: When the interaction between two variables changes depending on the level of a third variable.

c. Synergistic Interaction: When the combined effect of two variables is greater than the sum of their individual effects.

d. Antagonistic Interaction: When the effect of one variable diminishes the effect of another variable.

---

3. Visualizing Interaction Effects (15 mins)

Interaction plots:

- **Simple Interaction Plot:** A line graph showing how the levels of one independent variable change across the levels of another variable.
- **Contour plots or surface plots:** 3D visualizations used to explore more complex interactions.

---

4. Example of Interaction Effect (15 mins)

Scenario: Let's consider an example where we examine the effect of two factors—hours of study (Factor 1) and type of study material (Factor 2)—on student performance (dependent variable).

- **Factor 1:** Hours of Study (Low, Medium, High)
- **Factor 2:** Study Material Type (Textbook, Online)

If the effect of hours of study on performance depends on the type of study material used, there is an interaction effect between the two factors.

---

5. Python Code for Interaction Effects (30 mins)

Step 1: Import Libraries

python
Copy code

```
import pandas as pd
import numpy as np
import statsmodels.api as sm
import statsmodels.formula.api as smf
import matplotlib.pyplot as plt
import seaborn as sns
```

Step 2: Sample Data

We create a dataset simulating the example of study hours and study material.

python
Copy code

```
# Create sample data
data = {
    'study_hours': np.random.choice(['Low', 'Medium', 'High'],
size=100),
```

```
    'material_type': np.random.choice(['Textbook', 'Online'],
size=100),
    'performance': np.random.randint(50, 100, size=100)
}

df = pd.DataFrame(data)

# Convert categorical variables to dummy variables
df['study_hours_cat'] = df['study_hours'].astype('category').cat.codes
df['material_type_cat'] =
df['material_type'].astype('category').cat.codes
```

Step 3: Define the Interaction Model

Using an OLS model to estimate interaction effects.

python
Copy code

```
# Create an interaction term
df['interaction'] = df['study_hours_cat'] * df['material_type_cat']

# Build OLS regression model
model = smf.ols('performance ~ study_hours_cat * material_type_cat',
data=df).fit()
print(model.summary())
```

Step 4: Visualizing the Interaction Effect

python
Copy code

```
# Use Seaborn to visualize interaction effects
sns.pointplot(x='study_hours', y='performance', hue='material_type',
data=df)
plt.title("Interaction between Study Hours and Material Type on
Performance")
plt.show()
```

Explanation of Code:

- We first create dummy variables for the categorical independent variables.

- The interaction effect is modeled using `study_hours_cat * material_type_cat` to include the interaction term in the regression.
- Finally, a point plot helps visualize the interaction.

---

6. Interpreting Interaction Effects (15 mins)

Understanding the coefficients:

- Main effects (study_hours and material_type) show the effect of each variable when the other variable is held constant.
- The interaction term shows how the effect of one independent variable changes across the levels of the other independent variable.

Interpreting the plot:

- If the lines in the interaction plot are parallel, there is no interaction.
- If the lines cross or diverge, an interaction effect is present.

---

7. Practical Applications (15 mins)

1. Marketing: Analyzing the effect of price and advertising strategies on sales. The impact of a discount might depend on how the product is marketed.

2. Medicine: Examining how different drugs interact to affect patient health. One drug's efficacy might depend on whether another drug is also being taken.

3. Education: Studying how teaching style and class size together affect student achievement.

---

8. Q&A and Discussion (10 mins)

- Encourage students to ask questions.
- Ask students to come up with examples of interaction effects in their fields of interest.

Understanding Interaction Effects in Statistics

By Jim Frost

Ref: https://statisticsbyjim.com/regression/interaction-effects/

What are Interaction Effects?

An interaction effect occurs when the effect of one variable depends on the value of another variable.

Understanding Interaction Effects in Statistics

By Jim Frost 513 Comments

What are Interaction Effects?

An interaction effect occurs when the effect of one variable depends on the value of another variable. Interaction effects are common in regression models, ANOVA, and designed experiments. In this post, I explain interaction effects, the interaction effect test, how to interpret interaction models, and describe the problems you can face if you don't include them in your model.

In any study, whether it's a taste test or a manufacturing process, many variables can affect the outcome. Changing these variables can affect the outcome directly. For instance, changing the food condiment in a taste test can affect the overall enjoyment. In this manner, analysts use models to assess the relationship between each independent variable and the dependent variable.

This kind of an effect is called a main effect. While main effects are relatively straightforward, it can be a mistake to assess only main effects.

In more complex study areas, the independent variables might interact with each other. Interaction effects indicate that a third variable influences the relationship between an independent and dependent variable. In this situation, statisticians say that these variables interact because the relationship between an independent and dependent variable changes depending on the value of a third variable. This type of effect makes the model more complex, but if the real world behaves this way, it is critical to incorporate it in your model.

Example of Interaction Effects with Categorical Independent Variables

Imagine that we are conducting a taste test to determine which food condiment produces the highest enjoyment.

Our two independent variables are both categorical variables: Food and Condiment.

Factor Information

Factor	Type	Levels	Values
Food	Fixed	2	Hot Dog, Ice Cream
Condiment	Fixed	2	Chocolate Sauce, Mustard

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Food	1	1.6	1.6	0.06	0.801
Condiment	1	277.5	277.5	11.07	0.001
Food*Condiment	1	15695.8	15695.8	626.15	0.000
Error	76	1905.1	25.1		
Total	79	17880.0			

1. What are Interaction Effects?

- **Definition:**
  Interaction effects occur when the effect of one variable on the outcome depends

on the value of another variable. They are common in regression models, ANOVA, and designed experiments.

  - **Main effects** are the individual impact of each independent variable on the dependent variable.
  - **Interaction effects** indicate that the relationship between an independent variable and the dependent variable changes depending on another variable.
- **Why Interaction Effects Matter:**
  - Many variables may directly influence an outcome, and focusing solely on main effects can lead to incorrect conclusions.
  - In complex systems, independent variables often interact, and ignoring interaction effects can lead to misleading results.

---

2. Example of Interaction Effects (Categorical Variables)

- **Scenario:**
  A taste test is conducted with two foods (ice cream, hot dogs) and two condiments (chocolate sauce, mustard), assessing the dependent variable: satisfaction.
  - **Question:** Do you prefer ketchup or chocolate sauce on your food?
    The answer is: *It depends* on the type of food!
    This showcases an interaction effect: The effect of the condiment depends on the type of food.
- **ANOVA Model with Interaction Term:**
  Satisfaction=Food+Condiment+Food×Condiment\text{Satisfaction} = \text{Food} + \text{Condiment} + \text{Food} \times \text{Condiment}Satisfaction=Food+Condiment+Food×Condiment
  In this case, the interaction term (Food * Condiment) helps capture how enjoyment differs based on the food and condiment pairing.
- **Interpretation via Interaction Plot:**
  - In an interaction plot, parallel lines indicate no interaction.
  - If the lines cross or have different slopes, an interaction effect exists.
  - **Example Interpretation:**
    - Enjoyment is higher when chocolate sauce is on ice cream and mustard is on hot dogs.
    - If you mix these up, satisfaction drops (e.g., mustard on ice cream).

---

3. Visualizing and Testing Interaction Effects

- **Interaction Plots:**
  - X-axis: First independent variable (e.g., type of food).
  - Y-axis: Dependent variable (e.g., satisfaction).
  - Lines: Second independent variable (e.g., type of condiment).
- **Interpreting the Plot:**
  - **Parallel lines:** No interaction.
  - **Non-parallel or crossed lines:** Interaction effect is present.
- **P-values:**
  Interaction effects are tested using hypothesis tests. A significant p-value (e.g., < 0.05) for the interaction term confirms the presence of an interaction effect.

---

4. Dangers of Ignoring Interaction Effects

- Ignoring interaction effects can lead to incorrect conclusions.
  - **Example:**
    If we ignore the interaction between food and condiment, we might incorrectly choose hot dogs with chocolate sauce (which would lead to lower satisfaction), based solely on the main effects. Interaction effects give the *full* picture.

---

5. Interaction Effects with Continuous Variables

- **Scenario:**
  In a manufacturing process, the independent variables—processing time, temperature, and pressure—affect product strength. We include the interaction term between temperature and pressure in a regression model.
  - **Interaction Model:**
    Strength=Time+Temperature+Pressure+Temperature×Pressure\text{Strength} = \text{Time} + \text{Temperature} + \text{Pressure} + \text{Temperature} \times \text{Pressure}Strength=Time+Temperature+Pressure+Temperature×Pressure
- **Plot Interpretation:**
  - At high pressures, there is a positive relationship between temperature and strength.

    - At low pressures, the relationship becomes negative.
- This suggests that the optimal temperature for product strength depends on the pressure level—another "it depends" scenario.

---

6. Important Considerations

- **Statistical Significance:**
  While plots can help visualize interactions, use hypothesis tests (like p-values) to confirm the significance of interaction effects and rule out random sampling error.
- **Two-Way and Higher-Order Interactions:**
    - **Two-Way Interaction:** Involves two independent variables interacting (e.g., Food * Condiment).
    - **Three-Way Interaction:** Involves three variables (e.g., Food * Condiment * Time), though these are less common and harder to interpret.
- **Main Effects and Interaction Effects:**
  When significant interaction effects are present, main effects cannot be interpreted in isolation. Ignoring interactions leads to misleading interpretations.

---

7. Python Code for Interaction Effects

Step 1: Import Libraries

python

Copy code

```
import pandas as pd

import numpy as np

import statsmodels.api as sm

import statsmodels.formula.api as smf

import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

Step 2: Sample Data

python

Copy code

```
# Create sample data

data = {

    'food': np.random.choice(['Ice Cream', 'Hot Dog'],
size=100),

    'condiment': np.random.choice(['Chocolate Sauce',
'Mustard'], size=100),

    'satisfaction': np.random.randint(50, 100, size=100)

}



df = pd.DataFrame(data)



# Convert categorical variables to dummy variables

df['food_cat'] = df['food'].astype('category').cat.codes
```

```
df['condiment_cat'] =
df['condiment'].astype('category').cat.codes
```

Step 3: Define the Interaction Model

python

Copy code

```
# Create an interaction term

df['interaction'] = df['food_cat'] * df['condiment_cat']



# Build OLS regression model

model = smf.ols('satisfaction ~ food_cat * condiment_cat',
data=df).fit()

print(model.summary())
```

Step 4: Visualize Interaction Effects

python

Copy code

```
# Use Seaborn to visualize interaction effects

sns.pointplot(x='food', y='satisfaction', hue='condiment',
data=df)
```

```
plt.title("Interaction between Food and Condiment on Satisfaction")

plt.show()
```

---

8. Key Takeaways

- Interaction effects are critical in understanding how multiple variables jointly influence outcomes.
- Always check for interaction effects when analyzing data with multiple independent variables.
- Use visualization (interaction plots) and hypothesis testing to confirm the presence of interaction effects.

Interaction Effects with Numeric Predictors

---

1. Introduction to Interaction Effects with Numeric Predictors

- Interaction effects with numeric (continuous) predictors occur when the relationship between two numeric independent variables and the dependent variable changes based on the values of one of the predictors.
- In a regression model, interaction effects between continuous predictors help capture how one predictor's influence on the outcome is modified by another predictor's value.

---

2. Example Scenario: Manufacturing Process

- **Scenario:** We assess the impact of two numeric variables—**Temperature** and **Pressure**—on **Product Strength** in a manufacturing process.
  - **Objective:** Understand how product strength is affected by temperature and pressure, and whether the effect of temperature depends on pressure.
- **Regression Model with Interaction Term:**
  Strength=β0+β1Temperature+β2Pressure+β3(Temperature×Pressure)+ϵ\text{Strength}

= \beta_0 + \beta_1 \text{Temperature} + \beta_2 \text{Pressure} + \beta_3 (\text{Temperature} \times \text{Pressure}) + \epsilonStrength=β0+β1Temperature+β2Pressure+β3(Temperature×Pressure)+ϵ

- The interaction term Temperature×Pressure\text{Temperature} \times \text{Pressure}Temperature×Pressure captures how the effect of temperature on strength changes depending on the level of pressure.

---

3. Visualizing Interaction Effects

- **Interaction Plot for Numeric Predictors:**
  - X-axis: One of the independent variables (e.g., temperature).
  - Y-axis: Dependent variable (e.g., product strength).
  - Lines: Represent different levels of the other independent variable (e.g., pressure).
- In the plot, if the slopes of the lines differ (i.e., non-parallel lines), there is an interaction effect between the two variables.

---

4. Python Code for Interaction Effects with Numeric Predictors

Step 1: Import Required Libraries

python
Copy code

```python
import pandas as pd
import numpy as np
import statsmodels.api as sm
import statsmodels.formula.api as smf
import matplotlib.pyplot as plt
import seaborn as sns
```

Step 2: Generate Sample Data

python
Copy code

```python
# Generate sample data
np.random.seed(0)
n = 100
df = pd.DataFrame({
```

```
    'temperature': np.random.uniform(50, 150, n),  # Temperature
between 50 and 150 degrees
    'pressure': np.random.uniform(1, 5, n),        # Pressure between
1 and 5 units
    'strength': np.random.uniform(200, 500, n)     # Random product
strength values
})

# Adding an interaction effect between temperature and pressure on
product strength
df['strength'] = 200 + 3 * df['temperature'] - 5 * df['pressure'] +
0.5 * df['temperature'] * df['pressure'] + np.random.normal(0, 20, n)
```

Step 3: Build the Interaction Model

python
Copy code

```
# Create an interaction term between temperature and pressure
df['temp_pressure_interaction'] = df['temperature'] * df['pressure']

# Build the regression model with interaction term
model = smf.ols('strength ~ temperature * pressure', data=df).fit()
print(model.summary())
```

- The model summary will provide the p-value for the interaction term to test whether the effect of temperature on strength significantly depends on pressure.

Step 4: Visualize the Interaction Effect

python
Copy code

```
# Visualizing interaction between temperature and pressure
sns.lmplot(x='temperature', y='strength', hue='pressure', data=df,
scatter_kws={"s": 10}, palette="coolwarm", aspect=1.5)

# Adjust the plot title and labels
plt.title("Interaction between Temperature and Pressure on Product
Strength")
plt.xlabel("Temperature")
```

```
plt.ylabel("Product Strength")
plt.show()
```

- **Explanation:**
  - The **interaction plot** shows the relationship between temperature and product strength at different pressure levels.
  - The slopes of the lines vary for different levels of pressure, indicating an interaction effect.

---

5. Interpretation of Results

- **Coefficient for Interaction Term:**
  - A significant interaction term coefficient indicates that the effect of one predictor (e.g., temperature) on the outcome (e.g., strength) depends on the value of the other predictor (e.g., pressure).
- **Non-significant Interaction Term:**
  - If the interaction term is not significant, this implies that the relationship between temperature and strength does not depend on the level of pressure. The main effects alone are sufficient to describe the model.
- **Example of Interpretation:**
  - If the interaction term is significant and positive, it means that as pressure increases, the effect of temperature on product strength becomes more pronounced. For instance, at high pressure, increasing temperature may lead to a greater increase in strength compared to when pressure is low.

---

6. Key Points for Interaction Effects with Numeric Predictors

- Interaction effects are crucial for modeling real-world processes where relationships between variables change depending on other variables.
- Always check for interaction terms in models involving multiple continuous predictors.
- Visualizations such as interaction plots and hypothesis testing (p-values) help confirm the presence of interaction effects.