Advanced Data Visualization Lecture Notes (Lecture 9 - 2 Hours)

Interaction Effects with One Numeric and One Categorical Predictor

Duration: 2 hours

---

1. Introduction to Interaction Effects with Mixed Predictors

- Interaction effects can also occur between a numeric (continuous) predictor and a categorical predictor.
- In this case, the relationship between the numeric predictor and the dependent variable changes depending on the category of the categorical predictor.
- These models are common in fields like social science, marketing, and healthcare, where both numeric and categorical variables influence outcomes.

---

2. Example Scenario: Marketing Campaign

- **Scenario:** We examine how **Sales** (dependent variable) is affected by **Advertising Spend** (numeric predictor) and **Campaign Type** (categorical predictor with levels: "Online" and "Offline").
    - **Objective:** Understand whether the effect of advertising spend on sales differs depending on the campaign type.
- **Regression Model with Interaction Term:** $\text{Sales}=\beta_0+\beta_1\text{Advertising}+\beta_2\text{Campaign Type}+\beta_3(\text{Advertising}\times\text{Campaign Type})+\epsilon$\text{Sales} = \beta_0 + \beta_1 \text{Advertising} + \beta_2 \text{Campaign Type} + \beta_3 (\text{Advertising} \times \text{Campaign Type}) + \epsilon$\text{Sales}=\beta_0+\beta_1\text{Advertising}+\beta_2\text{Campaign Type}+\beta_3(\text{Advertising}\times\text{Campaign Type})+\epsilon$
    - The interaction term $\text{Advertising}\times\text{Campaign Type}$\text{Advertising} \times \text{Campaign Type}$\text{Advertising}\times\text{Campaign Type}$ captures whether the effect of advertising spend on sales depends on the type of campaign (online or offline).

---

3. Visualizing Interaction Effects

- **Interaction Plot for Mixed Predictors:**
    - X-axis: Numeric predictor (e.g., advertising spend).
    - Y-axis: Dependent variable (e.g., sales).

- Lines: Represent different categories of the categorical variable (e.g., online and offline campaign types).
- If the slopes of the lines differ (i.e., non-parallel lines), it indicates an interaction effect.

---

4. Python Code for Interaction Effects with One Numeric and One Categorical Predictor

Step 1: Import Required Libraries

python
Copy code

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

Step 2: Generate Sample Data

python
Copy code

```
# Create a sample dataset
df = pd.DataFrame({
    'advertising': [100, 150, 200, 250, 300, 100, 150, 200, 250, 300],
# Numeric predictor
    'campaign_type': ['Online', 'Online', 'Online', 'Online',
'Online', 'Offline', 'Offline', 'Offline', 'Offline', 'Offline'],  #
Categorical predictor
    'sales': [50, 60, 80, 90, 110, 30, 40, 50, 55, 70]  # Dependent
variable
})

# Convert the categorical variable to a format suitable for modeling
df['campaign_type'] = df['campaign_type'].astype('category')
```

Step 3: Build the Interaction Model

python
Copy code

```
# Build the interaction model using ordinary least squares (OLS)
regression
model = smf.ols('sales ~ advertising * campaign_type', data=df).fit()
print(model.summary())
```

- The model summary will display the coefficients for the main effects (advertising and campaign type) and the interaction term.

Step 4: Visualize the Interaction Effect

python
Copy code
```
# Visualize interaction effect
sns.lmplot(x='advertising', y='sales', hue='campaign_type', data=df,
aspect=1.5, palette="Set1")

# Add labels and title
plt.title('Interaction between Advertising and Campaign Type on
Sales')
plt.xlabel('Advertising Spend')
plt.ylabel('Sales')
plt.show()
```

- **Explanation:**
  - The **interaction plot** shows the relationship between advertising spend and sales for each campaign type (online and offline).
  - If the lines for "Online" and "Offline" have different slopes, this indicates that the effect of advertising spend on sales depends on the campaign type.

---

5. Interpretation of Results

- **Coefficients:**
  - The **interaction term** coefficient shows how the slope (relationship between advertising and sales) differs between the online and offline campaign types.
- **Significant Interaction Term:**
  - If the interaction term is significant, it means the effect of advertising on sales depends on the campaign type. For example, advertising spend may have a stronger effect on sales for online campaigns compared to offline campaigns.
- **Non-significant Interaction Term:**

    - If the interaction term is not significant, it means that the effect of advertising on sales is similar for both online and offline campaigns, and the main effects alone are sufficient to explain the relationship.

---

6. Key Points for Interaction Effects with Mixed Predictors

- **Interpreting Interaction Plots:**
    - In an interaction plot, non-parallel lines suggest an interaction effect between the numeric and categorical variables.
    - If the lines are parallel, this indicates that there is no interaction effect, meaning the numeric predictor affects the outcome similarly across categories.
- **Consideration of Main Effects:**
    - Like with any interaction model, do not interpret the main effects of the variables (e.g., advertising spend or campaign type) without considering the interaction term if it is statistically significant.

---

7. Example Interpretation

- If the slope for **online campaigns** is steeper than for **offline campaigns**, it suggests that **increasing advertising spend** has a larger effect on **sales** for online campaigns than for offline ones.
- If the interaction term is significant, you would conclude that **the effect of advertising depends on the campaign type**.