《DID 模型：从核心概念到应用分析》

Here is the English version of the review content, tailored to the key points of the exam (focus on understanding, application, and basic formulas, while ignoring complex derivations), and based on the provided document 12 Difference in Difference.pdf.

I. Core Idea and Basic Concepts of DID

1. **Essence of the Idea**
    - Compare the difference in outcomes between the **treated group (after intervention)** and the **control group (without intervention)** before and after the intervention, eliminating the influence of common trends to identify the true causal effect of the intervention.
    - **Core logic**: If the intervention is effective, the change in the treated group's outcome (post - pre) should significantly differ from that of the control group (post - pre).
2. **Key Terms**
    - **Treated Group**: The group receiving the policy intervention (e.g., the London community that changed its water source in John Snow's study).
    - **Control Group**: The group not receiving the intervention, used to simulate the "counterfactual " outcome of the treated group.
    - **Pre-period and Post-period**: Time phases for comparing outcome changes.

II. Basic Formulas and Models of DID

1. **Core Calculation Formula**
    - **DID estimator**:

$$\text{DID} = \left(\bar{Y}_{t,post} - \bar{Y}_{t,pre} \right) - \left(\bar{Y}_{c,post} - \bar{Y}_{c,pre} \right)$$

where $\bar{Y}_{t,post}$ is the mean outcome of the treated group post-intervention, $\bar{Y}_{t,pre}$ is the mean pre-intervention, and $\bar{Y}_{c,post}/\bar{Y}_{c,pre}$ are the corresponding means for the control group.

- **Interpretation**: The difference between the treated group's self-change and the control group 's self-change gives the net treatment effect after removing common trends.

2. **Regression Model Expression**
    - **Basic model**:

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 P_t + \beta_3 (D_i \times P_t) + \epsilon_{it}$$

where:

- $D_i = 1$ for the treated group, 0 for the control group;
- $P_t = 1$ for post-intervention, 0 for pre-intervention;
- The interaction term coefficient **is the DID estimator**, representing the net effect of the intervention on the treated group.

III. Key Assumptions and Tests

1. **Parallel Trends Assumption**
    - **Core condition**: Before the intervention, the outcome trends of the treated and control groups should be consistent (i.e., without intervention, their change trends are the same).
    - **Role**: Ensures the control group can effectively simulate the treated group's counterfactual outcome.
    - **Testing methods**:
        - **Graphical test**: Plot pre-intervention trend lines of both groups to observe parallelism.
        - **Statistical test**: Use multi-period pre-intervention data to test whether trend differences are significant (e.g., `estat ptrends` in Stata).
2. **Stable Unit Treatment Value Assumption (SUTVA)**
    - **Content**:
        - No spillover effects (intervention in the treated group does not affect the control group's outcomes);
        - Treatment is well-defined (no different treatment versions causing outcome variations).
    - **Example scenario**: When studying vaccine effects, assume vaccinated individuals do not influence unvaccinated individuals' infection risks (otherwise, SUTVA is violated).

IV. Classic Cases and Application Analysis

1. **John Snow's Cholera Study (Core Case in the Document)**
   - **Background**: In the 1850s, Snow aimed to verify if cholera was waterborne. One London community changed its water source (treated group), while others did not (control group).
   - **DID application**:
     - Treated group: Cholera death rate decreased post-water source change ($\bar{Y}_{t,post} - \bar{Y}_{t,pre}$);
     - Control group: Change in death rate of communities without water source change ($\bar{Y}_{c,post} - \bar{Y}_{c,pre}$);
     - **DID conclusion**: If the treated group's death rate decline is significantly greater, water is the primary infection source (ruling out air and other factors).
2. **South African Child Health Access Case (Empirical Case in the Document)**
   - **Question**: Evaluate the impact of community health clinic coverage on children's health (weight-for-age z-score, waz).
   - **Data and model**:
     - Treated group: 26 communities with clinic access in 1993;
     - Control group: 28 communities without access in 1993 (all gained access by 1998);
     - Regression model: `waz = β0 + β1 post + β2 hightreat + β3 post×hightreat + control variables`, where `post×hightreat` is the interaction term.
   - **Result**: The interaction term coefficient is 0.521 ($p<0.05$), indicating clinic coverage increased the treated group's waz by 0.521, a significant effect.

V. Steps to Build a DID Model (Key for Application)

1. **Define the research question and intervention scenario**
   - E.g., Evaluate the impact of a policy (e.g., Medicaid expansion) on hospital closure rates.
2. **Classify treated and control groups**
   - Treated group: Regions/groups affected by the policy;
   - Control group: Unaffected regions/groups with similar characteristics to the treated group (e.g., "nonexpansion states" in the Medicaid case).
3. **Define time phases**
   - Pre-intervention and post-intervention; multi-period data strengthens the parallel trends test.

4. **Construct the regression model**
    - Include treatment dummy, time dummy, their interaction term, and control for individual/group fixed effects and time fixed effects.
5. **Interpret the results**
    - The interaction term coefficient is the DID estimator; use significance to judge the intervention effect.
    - A positive and significant coefficient indicates the intervention improves the outcome, and vice versa.

VI. Summary of High-Frequency Exam Points

1. **Conceptual 辨析 (Conceptual Discrimination)**
    - Explain how DID removes confounding factors (e.g., common trends, fixed characteristics) via "double differencing".
    - Elucidate the meaning and importance of the parallel trends assumption.
2. **Formula Application**
    - Write the DID estimator expression and interpret each component.
    - Identify which coefficient in the regression model represents the DID effect.
3. **Case Analysis**
    - Design a DID model for a given scenario (e.g., the cholera or health access cases in the document): Define treated/control groups, time phases, write the regression equation, and explain result interpretation.
4. **Hypothesis Testing**
    - Outline methods to test the parallel trends assumption (graphical/statistical) and explain the impact of violating this assumption on results.

VII. Quick Memory Mnemonic

"Compare treated and control before and after; double difference removes trends.

The interaction term holds the effect; parallel trends are the premise.

Case modeling has three steps: group, time, and calculate the difference."