

《计量经济学中双重差分法（DID）高效复习》

以下是为你整理的计量经济学中“双重差分法（Difference-in-Differences, DID）”的高效复习内容，结合考试重点（重理解、重应用、轻推导）和文档案例，帮助你快速掌握核心考点。

一、DID 的核心思想与基本概念

1. 思想本质

- 通过比较**处理组（干预后）**和**控制组（无干预）**在**干预前后**的结果差异，剔除共同趋势的影响，从而识别干预的真实因果效应。
- 核心逻辑**：若干预有效，处理组的结果变化（后 - 前）应显著不同于控制组的结果变化（后 - 前）。

2. 关键术语

- 处理组（Treated Group）**：接受政策干预或处理的群体（如更换水源的伦敦社区）。
- 控制组（Control Group）**：未接受干预的群体（如未更换水源的社区），用于模拟处理组的“反事实”结果。
- 干预前（Pre-period）**和**干预后（Post-period）**：划分时间阶段，用于对比结果变化。

二、DID 的基本公式与模型

1. 核心计算公式

- DID 估计量**：

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

其中， $\bar{Y}_{t,\text{post}}$ 为处理组干预后均值， $\bar{Y}_{t,\text{pre}}$ 为处理组干预前均值， $\bar{Y}_{c,\text{post}}$ 和 $\bar{Y}_{c,\text{pre}}$ 为控制组对应时期均值。

- 含义**：处理组自身前后变化减去控制组自身前后变化，得到剔除共同趋势后的净处理效应。

2. 回归模型表达

- 基础模型**：

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

其中：

- $D_i = 1$ 为处理组，0 为控制组；
- $P_t = 1$ 为干预后，0 为干预前；
- 交互项系数** 即为 **DID 估计量**，表示处理组在干预后的净效应。

三、关键假设与检验

1. 平行趋势假设 (Parallel Trends Assumption)

- 核心条件：**干预前，处理组和控制组的结果趋势应一致（即若无干预，两组的变化趋势相同）。
- 作用：**确保控制组能有效模拟处理组的反事实结果。
- 检验方法：**
 - 图形检验：**绘制干预前两组结果的趋势图，观察是否平行。
 - 统计检验：**利用干预前多期数据，检验两组趋势差异是否显著（如 Stata 中的 `estat ptrends`）。

2. 稳定单位处理值假设 (SUTVA)

- 内容：**
 - 无干预溢出效应（处理组的干预不影响控制组的结果）；
 - 处理定义明确（无不同版本的处理导致结果差异）。
- 案例场景：**若研究疫苗效果，需假设接种者不影响未接种者的感染风险（否则违反 SUTVA）。

四、经典案例与应用分析

1. John Snow 与霍乱研究 (文档核心案例)

- 背景：**1850 年代，Snow 想验证霍乱是否通过水传播，伦敦某社区更换水源（处理组），其他社区未更换（控制组）。
- DID 应用：**
 - 处理组：更换水源后霍乱死亡率下降 ($\bar{Y}_{t,post} - \bar{Y}_{t,pre}$)；
 - 控制组：未更换水源的社区死亡率变化 ($\bar{Y}_{c,post} - \bar{Y}_{c,pre}$)；

- **DID 结论：**若处理组死亡率下降幅度显著大于控制组，说明水源是感染主因（排除空气等其他因素）。

2. 南非儿童健康访问案例（文档实证案例）

- **问题：**评估社区医疗诊所覆盖对儿童健康（体重年龄 Z 得分，waz）的影响。
- **数据与模型：**
 - 处理组：1993 年已覆盖诊所的 26 个社区；
 - 控制组：1993 年未覆盖、1998 年全覆盖的 28 个社区；
 - 回归模型： $waz = \beta_0 + \beta_1 post + \beta_2 hightreat + \beta_3 post \times hightreat + \text{控制变量}$ ，其中 $post \times hightreat$ 为交互项。
- **结果：**交互项系数为 0.521 ($p < 0.05$)，说明诊所覆盖使处理组儿童 waz 提高 0.521，效果显著。

五、DID 模型构建步骤（应用重点）

1. 确定研究问题与干预场景

- 例如：评估某政策（如医保扩张）对医院关闭率的影响。

2. 划分处理组与控制组

- 处理组：受政策影响的地区 / 群体；
- 控制组：未受政策影响、且与处理组具有相似特征的地区 / 群体（如文档中 Medicaid 扩张案例的“非扩张州”）。

3. 定义时间阶段

- 干预前（Pre）和干预后（Post），若有多期数据可增强平行趋势检验的可靠性。

4. 构建回归模型

- 包含处理组虚拟变量、时间虚拟变量及二者的交互项，控制个体 / 群体固定效应和时间固定效应。

5. 结果解释

- 交互项系数即为 DID 估计量，结合显著性判断干预效果；
- 若系数为正且显著，说明干预提升了结果变量，反之则降低。

六、考试高频考点总结

1. 概念辨析

- 解释 DID 如何通过“双重差分”剔除混淆因素（如共同趋势、固定特征）。

- 说明平行趋势假设的含义及为何重要。

2. 公式应用

- 写出 DID 估计量的表达式，并解释各部分含义。
- 指出回归模型中哪个系数代表 DID 效应。

3. 案例分析

- 根据给定场景（如文档中的霍乱、医疗覆盖案例），设计 DID 模型：确定处理组、控制组、时间阶段，写出回归方程，并说明如何解读结果。

4. 假设检验

- 简述平行趋势假设的检验方法（图形 / 统计），并说明若假设不成立对结果的影响。

七、快速记忆口诀

“处理控制比前后，双重差分去趋势；

交互项里藏效应，平行趋势是前提；

案例建模分三步，分组分时算差异。”

通过以上内容，重点理解 DID 的逻辑框架、假设条件和案例应用，结合回归模型中的交互项含义，即可快速应对考试中的概念题和应用题。

（注：文档部分内容可能由 AI 生成）