**Spatial Regression Discontinuity Design: School Fee Abolition Analysis**

**Exercise Overview**

In this exercise, you will learn to implement and interpret **spatial regression discontinuity design (RDD)** using Stata. You'll analyze the impact of a school fee abolition policy in the fictional Republic of Akwaaba (resembling Ghana) where the policy was implemented by district boundaries.

**Learning Objectives**

By the end of this exercise, you will be able to:

1. **Understand spatial RDD methodology** and when it's appropriate
2. **Create RDD plots** to visualize discontinuities at geographic boundaries
3. **Estimate treatment effects** using local linear regression with optimal bandwidth selection
4. **Incorporate spatial controls** to improve precision and address geographic confounders
5. **Conduct robustness checks** including balance tests and manipulation testing
6. **Interpret results** in the context of education policy in developing countries

**Policy Context**

**Setting**: Republic of Akwaaba, a fictional Sub-Saharan African country  
**Policy**: Free Primary Education Program implemented in Eastern Districts only  
**Research Question**: Did eliminating school fees increase enrollment, improve test scores, and reduce child marriage rates?

**Why Spatial RDD?** Geographic boundaries often create arbitrary treatment assignment - families on different sides of district boundaries are similar except for policy exposure, creating a natural experiment.

**Dataset Description**

You will work with data on **200 students** aged 6-18 from communities across the Republic of Akwaaba:

**Key Variables:**

* **Running Variable**: distance\_boundary - distance from the policy boundary (negative = west/control, positive = east/treatment)
* **Treatment**: fee\_abolition - whether student lives in district with free education policy
* **Outcomes**:
  + enrollment\_rate - school enrollment rate in student's community
  + test\_scores - standardized test scores (0-100 scale)
  + child\_marriage - child marriage indicator (females 14+ only)
* **Controls**: Household wealth, parental education, distance to school, urban/rural status, etc.
* **Spatial**: x\_coord, y\_coord geographic coordinates, dist\_from\_center distance from regional center

**Methodology: Standard vs Spatial RDD**

**Standard RDD:**

Uses only the running variable (distance from boundary) to estimate treatment effects around the cutoff.

**Spatial RDD Enhancement:**

Incorporates geographic controls (distance from center) to:

* Control for center-periphery economic gradients
* Improve precision by reducing residual variance
* Isolate boundary effects from general spatial patterns

**Key Analysis Steps**

**1. Visual Analysis**

Create RDD plots using rdplot to examine discontinuities in outcomes at the boundary.

**2. Estimation**

* Compare different bandwidth selections (manual vs MSE-optimal)
* Estimate both standard and spatial RDD specifications
* Use robust standard errors and community clustering

**3. Robustness Checks**

* **Balance tests**: Verify pre-treatment characteristics are smooth at boundary
* **Manipulation test**: Check for sorting around the boundary using McCrary density test
* **Sensitivity analysis**: Test stability across different specifications

**4. Interpretation**

Learn to interpret:

* Treatment effect magnitudes and statistical significance
* Confidence intervals and policy relevance
* Comparison between standard and spatial RDD results

**Expected Results**

Based on the data generation process, you should find:

* **Enrollment increases** by approximately 25 percentage points due to fee abolition
* **Test scores improve** by about 8 points on the 100-point scale
* **Child marriage rates decrease** by roughly 12 percentage points (for eligible females)
* **No manipulation** around the boundary (geographic boundaries are hard to sort around)
* **Balanced covariates** across the boundary (validating the RDD design)

**Technical Skills Developed**

**Stata Commands:**

* rdplot - Creating RDD visualization plots
* rdrobust - Optimal bandwidth selection and robust estimation
* rddensity - McCrary manipulation testing
* Working with spatial data and geographic controls

**Econometric Concepts:**

* Local polynomial estimation with triangular kernels
* Bias-corrected robust inference
* Bandwidth selection (MSE-optimal vs manual)
* Spatial correlation and appropriate clustering
* Balance testing and design validation

**Policy Implications**

This exercise demonstrates how spatial RDD can inform education policy:

**Immediate Impacts**: Quantify the direct effects of removing financial barriers to education

**Broader Effects**: Understand how education access affects long-term outcomes like child marriage

**Spatial Considerations**: Show how geographic context matters for policy effectiveness

**Cost-Benefit Analysis**: Provide evidence for scaling successful policies to other regions

**Real-World Applications**

Spatial RDD is valuable for analyzing:

* Administrative boundary-based policies (school districts, health zones)
* Geographic targeting of development programs
* Border discontinuities in regulations or services
* Electoral boundary effects on political outcomes

**Getting Started**

1. **Load the dataset**: use "spatial\_rdd\_education\_ghana.dta", clear
2. **Follow the do-file**: Work through each section systematically
3. **Pay attention to interpretation comments** - they explain what to look for in results
4. **Experiment with different specifications** - try the suggested variations
5. **Think critically about assumptions** - when might spatial RDD be inappropriate?

**Success Indicators**

You've mastered spatial RDD when you can:

* Identify appropriate settings for spatial RDD
* Create and interpret RDD plots showing clear discontinuities
* Estimate treatment effects with proper standard errors and clustering
* Validate design assumptions through balance and manipulation testing
* Compare spatial vs standard RDD and explain when geographic controls matter
* Communicate policy-relevant findings clearly

**Remember**: The goal isn't just to run commands, but to understand when and why spatial RDD provides credible causal identification. Focus on the economic intuition behind each test and what it tells us about the policy's effectiveness.

*Good luck with your analysis!*