

Lab9: Kriging

1. The different kriging models

วิธี Kriging อาศัยแนวคิดการออโตคอรีเลชัน (autocorrelation) โดยทั่วไปแล้ว ค่าความสัมพันธ์ (correlation) หมายถึงแนวโน้มของตัวแปรสองประเภทที่มีความเกี่ยวข้องกัน ตัวอย่างเช่น ตลาดหุ้นมักจะมีแนวโน้มเปลี่ยนแปลงในทิศทางตรงข้ามกับอัตราดอกเบี้ย ดังนั้นจึงกล่าวได้ว่าตลาดหุ้นมีความสัมพันธ์เชิงลบกับอัตราดอกเบี้ย อย่างไรก็ตาม ตลาดหุ้นมีออโตคอรีเลชันเชิงบวก หมายความว่าตลาดหุ้นมีความสัมพันธ์ในตัวเอง กล่าวคือ ราคาหุ้นในวันติดกันมักจะมีแนวโน้มคล้ายกันมากกว่าราคาหุ้นที่อยู่ห่างกันเป็นปี

แนวคิดนี้สอดคล้องกับหลักการพื้นฐานด้านธรณีวิทยาและภูมิศาสตร์ ซึ่งระบุว่าสิ่งที่อยู่ใกล้กันมักจะมีค่าคล้ายคลึงกันมากกว่าสิ่งที่อยู่ไกลกัน ระดับที่ความสัมพันธ์ลดลงสามารถแสดงเป็นฟังก์ชันของระยะทางได้ ออโตคอรีเลชันเป็นฟังก์ชันของระยะทาง ซึ่งเป็นลักษณะสำคัญของธรณีสถิติ (geostatistics) ในสถิติแบบดั้งเดิม การสังเกตจะถือว่าเป็นอิสระจากกัน นั่นคือ ไม่มีความสัมพันธ์ระหว่างค่าที่สังเกตได้ แต่ในธรณีสถิติ ข้อมูลเกี่ยวกับตำแหน่งเชิงพื้นที่ช่วยให้สามารถคำนวณระยะห่างระหว่างจุดสังเกตและสร้างแบบจำลองออโตคอรีเลชันเป็นฟังก์ชันของระยะทางได้

นอกจากนี้ เมื่อสังเกตว่าตลาดหุ้นโดยทั่วไปมีแนวโน้มเพิ่มขึ้นเมื่อเวลาผ่านไป ซึ่งเรียกว่าทิศทางแนวโน้ม (trend) สำหรับข้อมูลเชิงธรณีสถิติ คุณมีแนวคิดเดียวกัน และสามารถแสดงออกได้ในรูปของสมการทางคณิตศาสตร์ง่าย ๆ

$$Z(s) = \mu(s) + \varepsilon(s)$$

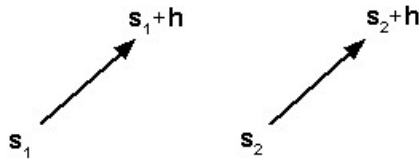
โดยที่ $Z(s)$ คือตัวแปรที่สนใจ สามารถถูกแยกออกเป็นสองส่วน ได้แก่ แนวโน้มหรือเทรนด์ $\mu(s)$ และข้อผิดพลาดแบบสุ่มที่มีออโตคอรีเลชัน $\varepsilon(s)$ สัญลักษณ์ s ใช้เพื่อแสดงตำแหน่งทางภูมิศาสตร์ ซึ่งสามารถคิดได้ว่าเป็นพิกัดเชิงพื้นที่ ได้แก่ ค่าพิกัด x (ลองจิจูด) และ y (ละติจูด)

รูปแบบต่างๆ ของสมการนี้เป็นพื้นฐานของวิธี Kriging ประเภทต่างๆ ในการทำความเข้าใจสมการนี้ ให้เริ่มจากด้านขวาแล้วค่อยๆ เคลื่อนมาทางซ้าย ไม่ว่ารูปแบบของเทรนด์ $\mu(s)$ ในโมเดลจะซับซ้อนแค่ไหน ก็ยังไม่สามารถทำนายค่าได้อย่างสมบูรณ์แบบ

ดังนั้นจึงต้องตั้งสมมติฐานเกี่ยวกับข้อผิดพลาด $\varepsilon(s)$ ได้แก่

- ค่าเฉลี่ยของข้อผิดพลาด $\varepsilon(s)$ ควรเป็นศูนย์ (โดยเฉลี่ย)
- ออโตคอรีเลชันระหว่าง $\varepsilon(s)$ และ $\varepsilon(s+h)$ ไม่ได้ขึ้นอยู่กับตำแหน่งจริง s แต่ขึ้นอยู่กระยะห่าง h ระหว่างสองจุดเท่านั้น

เงื่อนไขเหล่านี้จำเป็นเพื่อให้สามารถทำซ้ำได้และช่วยในการประมาณฟังก์ชันออโตคอรีเลชัน ตัวอย่างเช่น ในแผนภาพต่อไปนี้



ข้อผิดพลาดแบบสุ่มที่ตำแหน่งคู่ซึ่งเชื่อมโยงกันด้วยลูกศรถูกสมมติว่ามีอัตราค่าเฉลี่ยเท่ากัน หากพิจารณาเทรนด์ $\mu(s)$ เทรนด์นี้อาจเป็นค่าคงที่ง่ายๆ นั่นคือ $\mu(s) = m$ สำหรับทุกตำแหน่ง s และหาก m เป็นค่าที่ไม่ทราบ นี่คือลักษณะของโมเดลที่เป็นพื้นฐานของ **Ordinary Kriging** นอกจากนี้ เทรนด์ยังสามารถประกอบด้วยฟังก์ชันเชิงเส้นของพิกัดเชิงพื้นที่เองได้ เช่น

$$\mu(s) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy,$$

สมการที่กล่าวถึงเป็นพื้นผิวแนวโน้มนพหุนามลำดับสอง (second-order polynomial trend surface) ซึ่งเป็นเพียงการวิเคราะห์การถดถอยเชิงเส้น (linear regression) บนพิกัดเชิงพื้นที่ x และ y

Universal Kriging ใช้เมื่อแนวโน้ม $\mu(s)$ มีการเปลี่ยนแปลง และสัมประสิทธิ์ของการถดถอยไม่เป็นที่ทราบล่วงหน้า

Simple Kriging ใช้เมื่อแนวโน้มเป็นที่ทราบแน่นอน (ไม่ว่าจะเป็นค่าคงที่หรือไม่ก็ตาม) ซึ่งหมายถึงพารามิเตอร์และตัวแปรรวมทั้งหมดเป็นที่รู้จัก

จากสมการการแยกองค์ประกอบ $Z(s) = \mu(s) + \varepsilon(s)$ คุณสามารถทำการแปลง $Z(s)$ ได้ เช่น

- การเปลี่ยนให้เป็นตัวแปรตัวบ่งชี้ (indicator variable) ซึ่งกำหนดค่าเป็น 0 หาก $Z(s)$ ต่ำกว่าค่าที่กำหนด (เช่น 0.12 ppm สำหรับความเข้มข้นของโอโซน) และเป็น 1 หากสูงกว่าค่าดังกล่าว
- การพยากรณ์ความน่าจะเป็นที่ $Z(s)$ จะสูงกว่าค่าขั้นต่ำที่กำหนด โดยใช้ **Indicator Kriging**
- การแปลง $Z(s)$ ในลักษณะทั่วไปและกำหนดเป็น $f_i(Z(s_i))$ สำหรับตัวแปรตัวที่ i การสร้างตัวพยากรณ์ของฟังก์ชันของตัวแปร เช่น ถ้าคุณต้องการพยากรณ์ที่ตำแหน่ง s_0 คุณสามารถใช้ **Disjunctive Kriging** เพื่อนำค่าฟังก์ชันที่แปลงแล้วมาใช้

Cokriging: การใช้ตัวแปรหลายประเภท

ในกรณีที่มีมากกว่าหนึ่งตัวแปรที่สนใจ คุณสามารถสร้างโมเดลในรูปแบบ

$$Z_j(s) = \mu_j(s) + \varepsilon_j(s)$$

โดยที่ j คือดัชนีของตัวแปรแต่ละประเภท ซึ่งอาจมีแนวโน้มที่แตกต่างกัน นอกจากอัตราค่าเฉลี่ยของข้อผิดพลาด $\varepsilon_j(s)$ แล้ว ยังมี **Cross-correlation** ระหว่างข้อผิดพลาด $\varepsilon_j(s)$ และ $\varepsilon_k(s)$ ของตัวแปรสองประเภทอีกด้วย

ตัวอย่างเช่น ในการวิเคราะห์ข้อมูลสิ่งแวดล้อม คุณสามารถพิจารณาความสัมพันธ์ระหว่างความเข้มข้นของโอโซนและฝุ่นละอองขนาดเล็ก (PM) ซึ่งอาจไม่ได้ถูกวัดในตำแหน่งเดียวกันเสมอไป **Cokriging** เป็นแนวทางที่ใช้เมื่อมีมากกว่าหนึ่งตัวแปรที่เกี่ยวข้องกัน

โมเดล Cokriging สามารถขยายไปสู่หลายประเภท เช่น

- **Ordinary Cokriging**
- **Universal Cokriging**
- **Simple Cokriging**
- **Indicator Cokriging**
- **Probability Cokriging**
- **Disjunctive Cokriging**

ซึ่งเป็นการต่อยอดแนวคิดของ Kriging แบบต่างๆ ไปสู่กรณีที่มีหลายตัวแปร

Kriging in Geostatistical Analyst

Kriging ตั้งอยู่บนสมมติฐานที่ว่า อย่างน้อยบางส่วนของความแปรปรวนเชิงพื้นที่ที่พบในปรากฏการณ์ทางธรรมชาติสามารถจำลองได้โดยกระบวนการสุ่มที่มีอโตคอรืเลชันเชิงพื้นที่ และจำเป็นต้องสร้างแบบจำลองอโตคอรืเลชันนี้อย่างชัดเจน เทคนิค Kriging สามารถใช้เพื่อ

- อธิบายและจำลองรูปแบบเชิงพื้นที่
- ทำนายค่าที่ตำแหน่งที่ไม่ได้ทำการวัด
- ประเมินความไม่แน่นอนที่เกี่ยวข้องกับค่าที่ทำนาย

Geostatistical Wizard มีประเภทของ Kriging หลายแบบ ซึ่งเหมาะสมกับข้อมูลประเภทต่างๆ และมีสมมติฐานที่แตกต่างกัน ได้แก่

- **Ordinary Kriging**
- **Simple Kriging**
- **Universal Kriging**
- **Indicator Kriging**
- **Probability Kriging**
- **Disjunctive Kriging**
- **Empirical Bayesian Kriging**
- **Areal Interpolation**

ประเภทของแผนที่ที่สามารถสร้างได้จาก Kriging

วิธีการเหล่านี้สามารถใช้เพื่อสร้างพื้นผิวต่อไปนี้

- แผนที่ของค่าที่ทำนายโดย Kriging
- แผนที่ของค่าความผิดพลาดมาตรฐาน (Standard Errors) ที่เกี่ยวข้องกับค่าที่ทำนาย
- แผนที่ของความน่าจะเป็น ที่ระบุว่าค่าที่ทำนายเกินระดับวิกฤตที่กำหนดไว้หรือไม่
- แผนที่ของควอนไทล์ (Quantiles) สำหรับระดับความน่าจะเป็นที่กำหนดไว้ล่วงหน้า

ข้อยกเว้น:

1. **Indicator และ Probability Kriging** สามารถสร้างแผนที่ต่อไปนี้
 - แผนที่ของความน่าจะเป็น ที่ระบุว่าค่าที่ทำนายเกินระดับวิกฤตที่กำหนดไว้หรือไม่

- แผนที่ของค่าความผิดพลาดมาตรฐานของตัวบ่งชี้ (Indicator Standard Errors)
2. **Areal Interpolation** สามารถสร้างแผนที่ต่อไปนี้
- แผนที่ของค่าที่ทำนาย
 - แผนที่ของค่าความผิดพลาดมาตรฐานที่เกี่ยวข้องกับค่าที่ทำนาย

องค์ประกอบสำคัญของโมเดล Geostatistical

การสร้างแบบจำลองทางธรณีสถิติมีองค์ประกอบสำคัญหลายอย่าง ได้แก่

1. การวิเคราะห์ข้อมูลเชิงสำรวจเชิงพื้นที่ (ESDA) และ Variography เพื่อทำความเข้าใจโครงสร้างของข้อมูล
2. การสร้างแบบจำลอง Kriging ให้เหมาะสมกับความต้องการของคุณ
3. การตรวจสอบความถูกต้องของผลลัพธ์ ผ่าน Cross-validation และ Validation และเปรียบเทียบแบบจำลองต่างๆ เพื่อเลือกแบบที่ดีที่สุด

2. Ordinary kriging

Ordinary Kriging ตั้งอยู่บนสมมติฐานของโมเดล

$$Z(s) = \mu + \varepsilon(s)$$

โดยที่ μ เป็นค่าคงที่ที่ไม่ทราบล่วงหน้า

ข้อพิจารณาของ Ordinary Kriging

หนึ่งในประเด็นหลักของ **Ordinary Kriging** คือ สมมติฐานที่ว่าค่าเฉลี่ยเป็นค่าคงที่นั้น สมเหตุสมผลหรือไม่

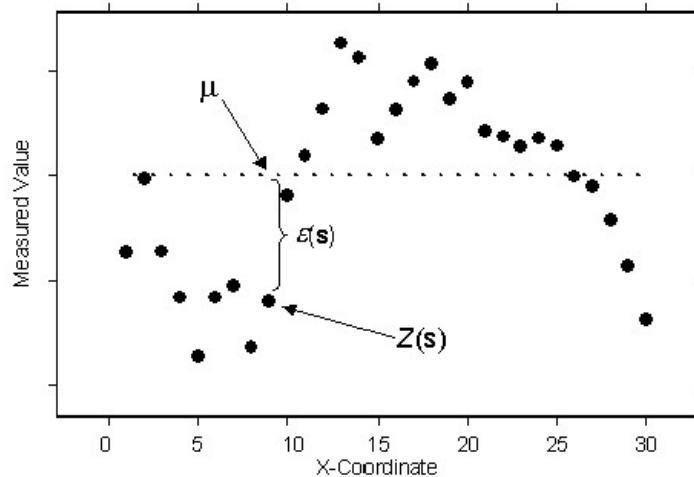
- ในบางกรณี อาจมีเหตุผลทางวิทยาศาสตร์ที่ทำให้ต้องปฏิเสธสมมติฐานนี้
- อย่างไรก็ตาม ในฐานะที่เป็นวิธีการพยากรณ์ที่เรียบง่าย **Ordinary Kriging** มีความยืดหยุ่นสูงและสามารถใช้งานได้ดีในหลายสถานการณ์

แผนภาพต่อไปนี้แสดงตัวอย่างของ Ordinary Kriging ในมิติเดียวเชิงพื้นที่ ซึ่งช่วยให้เข้าใจการพยากรณ์ค่าที่ตำแหน่งที่ไม่ได้วัดโดยใช้สมมติฐานของค่าเฉลี่ยคงที่ ตัวอย่างข้อมูลที่แสดงดูเหมือนเป็น **ค่าความสูง (Elevation)** ที่ถูกรวบรวมจาก **เส้นตัดผ่านหุบเขาและภูเขา**

- ด้านซ้ายของแผนภาพแสดงความแปรปรวนของข้อมูลที่มากกว่า
- ด้านขวาของแผนภาพดูเรียบขึ้น

ในความเป็นจริง ข้อมูลนี้ถูกจำลองขึ้นโดยใช้ **Ordinary Kriging Model** ที่มีค่าเฉลี่ยคงที่ μ

- เส้นประแสดงค่าเฉลี่ยที่แท้จริง (แต่ไม่ทราบล่วงหน้า)



ความสามารถของ Ordinary Kriging

1. สามารถใช้ได้ทั้ง **Semivariogram** และ **Covariance Function** เพื่อแสดงอโตคอร์เรลชัน
2. สามารถ แปลงข้อมูล (**Transformations**) และลบแนวโน้ม (**Remove Trends**) เพื่อให้โมเดลเหมาะสมขึ้น
3. สามารถรองรับ ค่าความผิดพลาดในการวัด (**Measurement Error**) เพื่อเพิ่มความแม่นยำในการพยากรณ์

3.Universal Kriging

Universal Kriging ตั้งอยู่บนสมมติฐานของโมเดล

$$Z(s) = \mu(s) + \varepsilon(s)$$

โดยที่ $\mu(s)$ เป็นฟังก์ชันเชิงกำหนด (Deterministic Function) ซึ่งอาจเปลี่ยนแปลงไปตามตำแหน่ง s

ตัวอย่างและการเปรียบเทียบกับ Ordinary Kriging

- แผนภาพที่แสดงใช้ ข้อมูลเดียวกัน กับที่ใช้ในแนวคิดของ **Ordinary Kriging**
- จุดทึบ (**Solid Circles**) ในแผนภาพแสดงค่าที่สังเกตได้

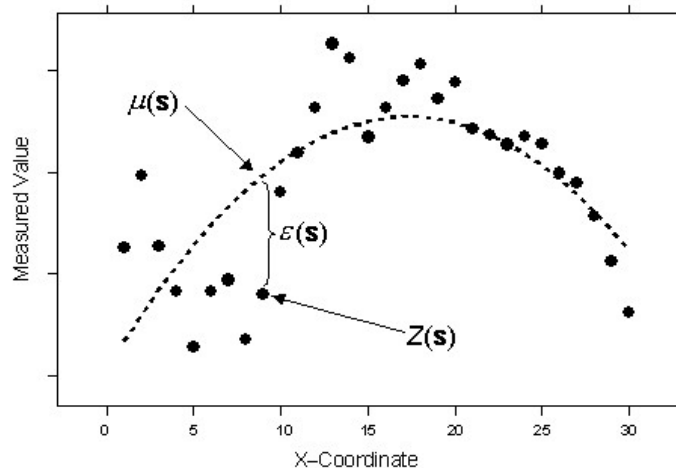
ข้อแตกต่างจาก Ordinary Kriging

- ใน **Ordinary Kriging** ค่าเฉลี่ย μ ถือเป็นค่าคงที่ที่ไม่ทราบล่วงหน้า
- ใน **Universal Kriging** ค่าเฉลี่ย $\mu(s)$ ไม่จำเป็นต้องเป็นค่าคงที่ แต่สามารถเปลี่ยนแปลงไปตามตำแหน่ง s ได้

Universal Kriging มักใช้ในกรณีที่ข้อมูลแสดงแนวโน้มที่ชัดเจน ซึ่งหมายความว่าอาจมีฟังก์ชันกำหนดแนวโน้ม (Trend Function) อยู่ เช่น

- แนวโน้มเชิงเส้น (Linear Trend)
- แนวโน้มพหุนามลำดับที่สอง (Second-Order Polynomial Trend)

Universal Kriging จึงเป็นเทคนิคที่เหมาะสมสำหรับข้อมูลที่ไม่สามารถใช้สมมติฐานของค่าเฉลี่ยคงที่ได้



ข้อควรระวังและข้อจำกัด

- เช่นเดียวกับ **Ordinary Kriging** ไม่มีวิธีที่แน่นอนในการตัดสินใจเลือกข้อมูลเพียงอย่างเดียวว่า แนวโน้มที่สังเกตได้นั้นเกิดจากแนวโน้มหรือออโตคอร์เรลชันของข้อผิดพลาด
- จำเป็นต้องใช้ ความรู้ทางวิทยาศาสตร์ หรือ การทดสอบทางสถิติ เพื่อช่วยระบุแนวโน้มที่เหมาะสม

คุณสมบัติของ Universal Kriging

- สามารถใช้ **Semivariogram** หรือ **Covariance Function** เพื่อจำลองออโตคอร์เรลชัน
- สามารถใช้ การแปลงข้อมูล (Transformations) เพื่อให้ข้อมูลเหมาะสมขึ้น
- สามารถรองรับ ข้อผิดพลาดในการวัด (Measurement Error)

Universal Kriging เป็นวิธีที่มีความยืดหยุ่นสูง โดยเฉพาะเมื่อข้อมูลแสดงแนวโน้มที่เปลี่ยนแปลงไปตามตำแหน่ง แต่ยังคงมีออโตคอร์เรลชันในข้อผิดพลาด

4.Simple Kriging

Simple Kriging ตั้งอยู่บนสมมติฐานของโมเดล

$Z(s) = \mu + \varepsilon(s)$ โดยที่:

- μ เป็น ค่าคงที่ที่ทราบล่วงหน้า (Known Constant)
- $\varepsilon(s)$ เป็นข้อผิดพลาดแบบสุ่มที่มีค่าเฉลี่ยเป็นศูนย์และมีออโตคอร์เรลชัน

ความแตกต่างของ Simple Kriging กับ Ordinary และ Universal Kriging

1. **Simple Kriging** → μ เป็นค่าที่ทราบแน่นอน (**Fixed & Known Mean**)
2. **Ordinary Kriging** → μ เป็นค่าคงที่ที่ **ไม่ทราบล่วงหน้า** และต้องประมาณค่า (**Unknown but Constant Mean**)
3. **Universal Kriging** → $\mu(s)$ เปลี่ยนแปลงตามตำแหน่ง s และต้องประมาณค่า (**Spatially Varying Mean**)

ตัวอย่างและแผนภาพ

- ใช้ข้อมูลชุดเดียวกันกับ **Ordinary Kriging** และ **Universal Kriging**
- จุดทึบ (**Solid Circles**) ในแผนภาพแสดงค่าที่สังเกตได้

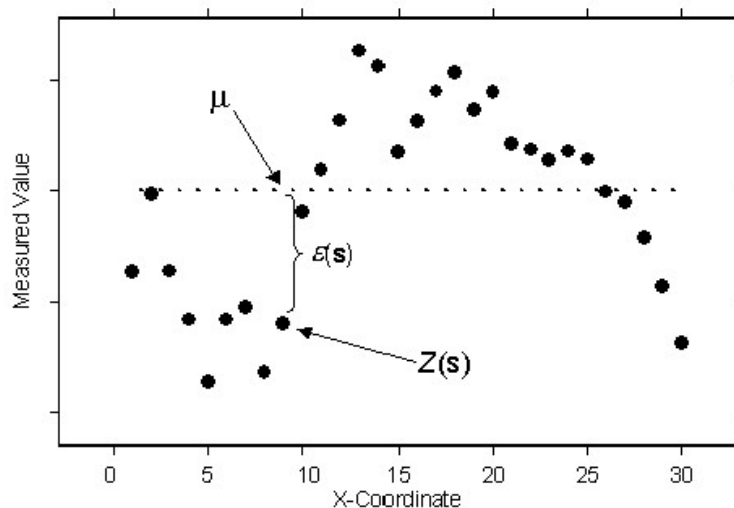
ข้อดีของ Simple Kriging

- ง่ายต่อการคำนวณ เนื่องจาก μ เป็นค่าคงที่ที่ทราบแล้ว
- ให้ผลลัพธ์ที่ดีในกรณีที่ข้อมูลมีค่าเฉลี่ยที่คงที่ทั่วพื้นที่

ข้อจำกัดของ Simple Kriging

- อาจไม่เหมาะสมหากค่าเฉลี่ยของข้อมูลเปลี่ยนแปลงไปตามตำแหน่ง
- ไม่สามารถจับแนวโน้ม (Trend) ของข้อมูลได้ดีเท่า **Universal Kriging**

Simple Kriging เป็นวิธีที่มีข้อกำหนดเข้มงวดกว่าวิธีอื่น เนื่องจากต้องรู้ค่าเฉลี่ยล่วงหน้า แต่ถ้าสมมติฐานนี้ถูกต้อง ก็จะทำให้ผลลัพธ์ที่ดีที่สุด



ค่าคงที่ที่ทราบค่า (known constant) ซึ่งแสดงด้วยเส้นประ คือ μ สามารถเปรียบเทียบกับ **ordinary kriging** ได้

สำหรับ simple kriging

- ถือว่า μ เป็นค่าที่ทราบแน่นอน
- ดังนั้น จะทราบ $\varepsilon(s)$ อย่างแน่นอน ที่ตำแหน่งของข้อมูล

แต่สำหรับ **ordinary kriging**

- μ ต้องถูกประมาณค่า (estimated)
- ดังนั้น $\varepsilon(s)$ ก็ต้องถูกประมาณค่าเช่นกัน

หากเราทราบค่า $\varepsilon(s)$ เราจะสามารถทำ การประมาณค่าความสัมพันธ์กันเองตามตำแหน่ง (autocorrelation) ได้แม่นยำขึ้น เมื่อเทียบกับกรณีที่ต้อง ประมาณค่า $\varepsilon(s)$ ไปพร้อมกับ การพยากรณ์

อย่างไรก็ตาม สมมติฐานที่ว่าค่าเฉลี่ย (mean, μ) เป็นค่าที่ทราบแน่นอนมักไม่สมจริง แต่ในบางกรณี การใช้ แบบจำลองที่อิงตามหลักฟิสิกส์ อาจให้แนวโน้ม (trend) ที่ทราบค่า ได้ ดังนั้น สามารถนำความแตกต่างระหว่างแบบจำลองกับค่าที่สังเกตได้ (เรียกว่า residuals) แล้วใช้ simple kriging กับ residuals ได้ โดยถือว่าสมการแนวโน้มของ residuals มีค่าเป็นศูนย์

Simple kriging สามารถ

- ใช้ semivariograms หรือ covariances (แสดง autocorrelation)
- ใช้ การแปลงข้อมูล (transformations)
- รองรับ ข้อผิดพลาดจากการวัด (measurement error)

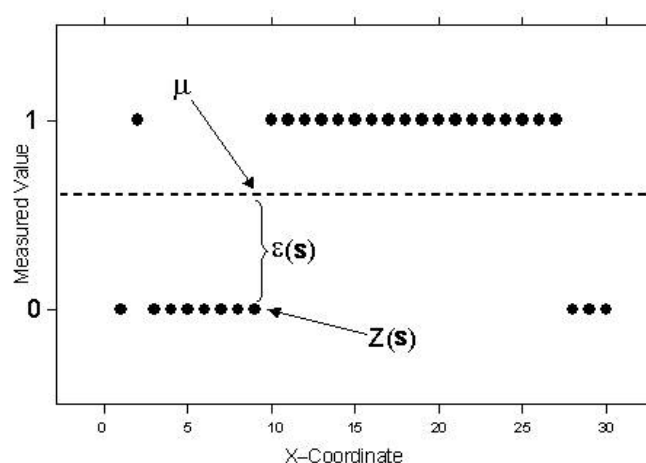
5.Indicator Kriging

Indicator Kriging ตั้งอยู่บนสมมติฐานของโมเดล

$$I(s) = \mu + \varepsilon(s)$$

โดยที่ μ เป็นค่าคงที่ที่ไม่ทราบค่า และ $I(s)$ เป็นตัวแปรเลขฐานสอง (binary variable) การสร้างข้อมูลเลขฐานสองอาจทำได้โดยใช้เกณฑ์ (threshold) กับข้อมูลเชิงต่อเนื่อง หรืออาจเป็นเพราะข้อมูลที่สังเกตได้มีค่าเป็น 0 หรือ 1 อยู่แล้ว ตัวอย่างเช่น คุณอาจมีตัวอย่างข้อมูลที่แสดงว่าจุดหนึ่งเป็นป่าหรือไม่ใช่ป่า โดยที่ตัวแปรเลขฐานสองใช้ระบุประเภทของพื้นที่นั้น

เมื่อใช้ตัวแปรเลขฐานสอง วิธี Indicator Kriging จะดำเนินการเช่นเดียวกับ Ordinary Kriging ในรูปต่อไปนี่ ข้อมูลถูกแปลงเป็นค่าทางเลขฐานสองโดยใช้เกณฑ์ที่แสดงในหัวข้อ **Understanding thresholds**



ข้อมูลเลขฐานสองที่สังเกตได้จะแสดงด้วยสี่เหลี่ยมเปล่า (open squares) ค่าเฉลี่ยที่ไม่ทราบค่าของตัวแปรตัวชี้วัดทั้งหมด (indicator variables) แสดงด้วยเส้นประ ซึ่งก็คือค่า μ

สามารถเปรียบเทียบสิ่งนี้กับ **ordinary kriging** ได้ โดยเช่นเดียวกับ ordinary kriging จะถือว่า $\epsilon(s)$ มีความสัมพันธ์กันเองตามตำแหน่ง (**autocorrelated**) เนื่องจากตัวแปรตัวชี้วัดมีค่าเป็น **0** หรือ **1** การประมาณค่า (interpolation) ที่ได้จะอยู่ระหว่าง 0 และ 1

ดังนั้น การพยากรณ์ (prediction) จาก **indicator kriging** สามารถตีความได้ว่าเป็น **ความน่าจะเป็นของตัวแปรที่จะมีค่าเป็น 1** หรือเป็นความน่าจะเป็นที่จุดนั้นจะอยู่ในประเภทที่กำหนดโดยค่า **1**

หากมีการใช้เกณฑ์ (threshold) เพื่อสร้างตัวแปรตัวชี้วัด แผนที่ที่ได้จากการประมาณค่า (interpolation map) จะแสดง **ความน่าจะเป็นของค่าที่สูงกว่า (หรือค่าที่ต่ำกว่า) เกณฑ์ที่กำหนด**

สามารถสร้างตัวแปรตัวชี้วัดหลายตัวจากชุดข้อมูลเดียวกันได้ โดยใช้หลายเกณฑ์ (**multiple thresholds**) ในกรณีนี้

- เกณฑ์หลักจะสร้างตัวแปรตัวชี้วัดหลัก
- ตัวแปรตัวชี้วัดอื่น ๆ จะถูกใช้เป็น **ตัวแปรรอง** ใน **cokriging**

Indicator kriging สามารถใช้ **semivariograms** หรือ **covariances** ซึ่งเป็นรูปแบบทางคณิตศาสตร์ที่ใช้แสดง **ความสัมพันธ์กันเองตามตำแหน่ง (autocorrelation)**

6. Probability Kriging

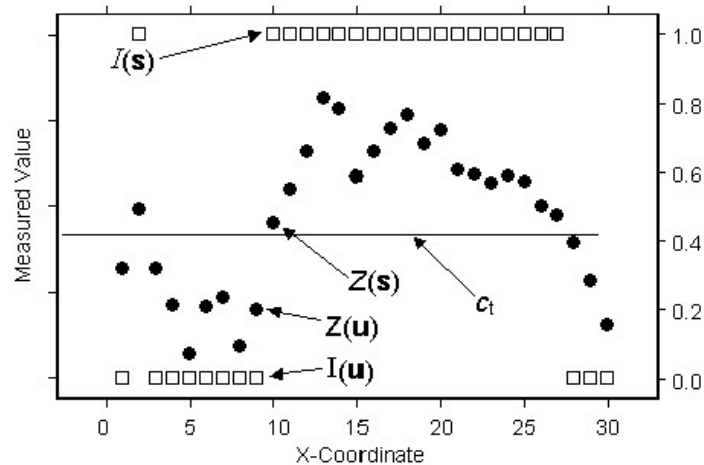
Probability Kriging ตั้งอยู่บนสมมติฐานของโมเดล

$$I(s) = I(Z(s) > ct) = \mu_1 + \epsilon_1(s)$$
$$Z(s) = \mu_2 + \epsilon_2(s)$$

โดยที่ μ_1 และ μ_2 เป็นค่าคงที่ที่ไม่ทราบค่า และ **$I(s)$** เป็นตัวแปรเลขฐานสอง ซึ่งถูกสร้างขึ้นโดยใช้ตัวชี้วัดเกณฑ์ (threshold indicator) ในรูปแบบ **$I(Z(s) > ct)$**

สังเกตว่า ตอนนี้มี **ข้อผิดพลาดแบบสมสองประเภท** คือ $\epsilon_1(s)$ และ $\epsilon_2(s)$ ดังนั้น จะมีทั้ง **autocorrelation** สำหรับแต่ละตัวและ **ความสัมพันธ์ cross-correlation** ระหว่างทั้งสองตัว

Probability kriging มีจุดมุ่งหมายเดียวกันกับ **indicator kriging** แต่ใช้ **cokriging** เพื่อพยายามให้ได้ผลลัพธ์ที่ดียิ่งขึ้น ตัวอย่างเช่น ในรูปด้านล่าง ซึ่งใช้ชุดข้อมูลเดียวกันกับแนวคิดของ **ordinary, universal, simple** และ **indicator kriging** สังเกตจุดข้อมูลที่ระบุเป็น **$Z(u=9)$** ซึ่งมีตัวแปรตัวชี้วัด **$I(u) = 0$** และ **$Z(s=10)$** ซึ่งมีตัวแปรตัวชี้วัด **$I(s) = 1$**



หากต้องการพยากรณ์ค่าที่ตำแหน่งกึ่งกลางระหว่างจุดข้อมูลทั้งสอง (ที่พิกัด $x = 9.5$), การใช้ **indicator kriging** เพียงอย่างเดียว จะให้ค่าพยากรณ์ใกล้เคียงกับ **0.5**

อย่างไรก็ตาม สังเกตว่า

- **Z(s)** มีค่าที่สูงกว่าเกณฑ์เพียงเล็กน้อย
- **Z(u)** มีค่าต่ำกว่าเกณฑ์มาก

ดังนั้น จึงมีเหตุผลที่จะเชื่อว่า ค่าพยากรณ์ตัวชี้วัด (**indicator prediction**) ที่ตำแหน่ง **9.5** ควรน้อยกว่า **0.5**

Probability kriging พยายามใช้ข้อมูลเพิ่มเติมจากชุดข้อมูลดั้งเดิม นอกเหนือจากตัวแปรเลขฐานสอง แต่ก็มีต้นทุนที่ต้องจ่าย นั่นคือ

- ต้องทำ การประมาณค่ามากขึ้น
- รวมถึงต้อง ประมาณค่าความสัมพันธ์กันเองตามตำแหน่ง (**autocorrelation**) ของแต่ละตัวแปร
- และ ความสัมพันธ์ข้ามกัน (**cross-correlation**) ระหว่างตัวแปร

ยิ่งต้องประมาณค่าพารามิเตอร์ความสัมพันธ์ที่ไม่ทราบค่ามากขึ้นเท่าใด ก็ยิ่งเพิ่ม ความไม่แน่นอน (**uncertainty**) ดังนั้น **probability kriging** อาจไม่คุ้มค่ากับความพยายามที่เพิ่มขึ้น

Probability kriging สามารถใช้

- **semivariograms** หรือ **covariances** (ใช้แสดง autocorrelation)
- **cross-covariances** (ใช้แสดง cross-correlation)
- การแปลงข้อมูล (**transformations**)

แต่ ไม่สามารถรองรับข้อผิดพลาดจากการวัด (**measurement error**) ได้

7. Comparing models

การเปรียบเทียบโมเดล ใน Geostatistical Analyst เป็นขั้นตอนที่สำคัญในการประเมินว่าโมเดลที่ใช้สร้าง Geostatistical Layer นั้นมีความแม่นยำและมีประสิทธิภาพแค่ไหนเมื่อเทียบกับโมเดลอื่นๆ โดยสามารถใช้เปรียบเทียบโมเดลสองตัวที่ถูกสร้างขึ้นจากวิธีการ Interpolation ที่แตกต่างกัน (เช่น **IDW** กับ **Ordinary Kriging**) หรือใช้วิธีการเดียวกันแต่ใช้พารามิเตอร์ต่างกัน

ประเภทของการเปรียบเทียบ

1. การเปรียบเทียบวิธีการต่างๆ เช่น การเปรียบเทียบระหว่าง **IDW (Inverse Distance Weighted)** กับ **Ordinary Kriging** ว่าวิธีไหนเหมาะสมที่สุดสำหรับชุดข้อมูลของคุณ
2. การเปรียบเทียบพารามิเตอร์ของวิธีเดียวกัน ในกรณีที่ใช้วิธีเดียวกัน เช่น **Kriging** แต่ใช้พารามิเตอร์ที่ต่างกัน เช่น **Semivariogram Model** หรือ **Search Neighborhood** การเปรียบเทียบนี้จะช่วยให้เห็นผลกระทบของการเปลี่ยนแปลงพารามิเตอร์ต่อผลลัพธ์ของการพยากรณ์

การใช้ Comparison Dialog Box สำหรับเปรียบเทียบโมเดล

ใน **Geostatistical Analyst** จะมี **Comparison Dialog Box** ซึ่งใช้สำหรับเปรียบเทียบโมเดลต่างๆ โดยใช้สถิติเกี่ยวกับ **cross-validation** ที่ได้พูดถึงไปแล้วในการทำ **Cross-validation** และ **Validation**

เครื่องมือนี้ช่วยให้คุณสามารถดู สถิติ และ กราฟ ของโมเดลสองตัวพร้อมกันเพื่อให้การตัดสินใจในการเลือกโมเดลที่ดีที่สุดทำได้ง่ายขึ้น

เกณฑ์การเลือกโมเดลที่ดีที่สุด

โดยทั่วไปแล้ว โมเดลที่ดีที่สุด คือโมเดลที่มี:

1. **Mean** ที่เป็นมาตรฐานใกล้เคียงกับศูนย์
2. **Root-Mean-Squared Prediction Error** ที่น้อยที่สุด
3. ค่าเฉลี่ยของ **Standard Error** ใกล้เคียงกับ **Root-Mean-Squared Prediction Error**
4. **Standardized Root-Mean-Squared Prediction Error** ใกล้เคียงกับ 1

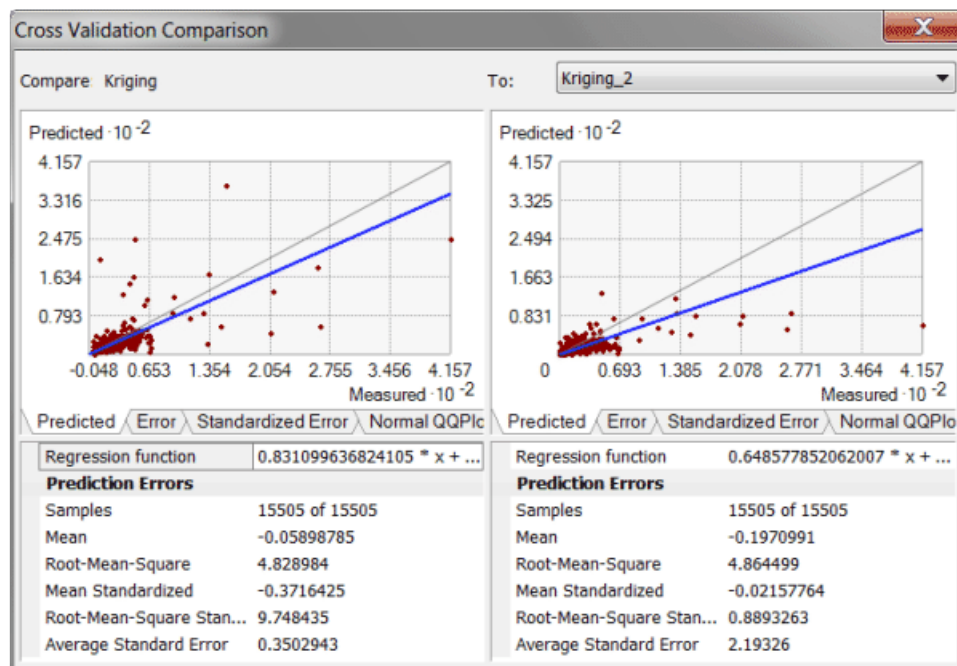
ขั้นตอนการเปรียบเทียบโมเดล

- ปกติแล้วจะสร้าง **Interpolate Surfaces** หลายโมเดลก่อนที่จะระบุว่าโมเดลไหนดีที่สุด ซึ่งโมเดลที่ดีที่สุดจะถูกใช้เป็นโมเดลสุดท้าย หรือจะถูกนำไปใช้ในโมเดลใหญ่ขึ้น เช่น **Suitability Model** สำหรับการกำหนดสถานที่สร้างบ้าน หรือการแก้ปัญหาที่มีอยู่
- สามารถเปรียบเทียบแต่โมเดล แล้วกำจัด โมเดลที่แย่ที่สุด จนกว่าคุณจะเหลือสองพื้นผิวที่ดีที่สุด และเปรียบเทียบกันอีกครั้ง

การใช้ Comparison Dialog Box:

- เปรียบเทียบโมเดลต่างๆ ได้อย่างเป็นระบบ
- เลือกโมเดลที่ดีที่สุดตามสถิติและกราฟที่แสดงผล
- สร้างโมเดลที่มีความแม่นยำสูงสุดสำหรับใช้งานในงานต่างๆ เช่น การจำลองสภาพเหมาะสมของพื้นที่สำหรับการตั้งบ้าน

การเปรียบเทียบเหล่านี้ทำให้การเลือก **Geostatistical Surface** ที่ดีที่สุดเพื่อแก้ปัญหาหรือการวิเคราะห์เป็นไปอย่างมีประสิทธิภาพและเหมาะสมที่สุด!



ข้อควรพิจารณาเมื่อเปรียบเทียบวิธีการและโมเดล

เมื่อคุณเปรียบเทียบผลลัพธ์จาก วิธีการ หรือ โมเดล ต่างๆ มีสองประเด็นที่สำคัญที่ต้องพิจารณา:

1. **Optimality** (ความเหมาะสมที่สุด)
2. **Validity** (ความถูกต้อง)

Optimality (ความเหมาะสมที่สุด)

- ตัวอย่างเช่น **Root-Mean-Squared Prediction Error** อาจมีค่าน้อยในโมเดลหนึ่งซึ่งอาจทำให้คุณสรุปว่าโมเดลนี้ดีที่สุด
- แต่เมื่อลองเปรียบเทียบกับโมเดลอีกตัวที่มี **Root-Mean-Squared Prediction Error** ที่ใกล้เคียงกับ **Average Estimated Prediction Standard Error** โมเดลนั้นอาจจะมีค่าถูกต้องมากกว่า เนื่องจากเมื่อพยากรณ์ ณ จุดที่ไม่มีข้อมูล เราจะใช้ **Standard Errors** ที่ประมาณการ ในการประเมินความไม่แน่นอนของการพยากรณ์นั้น

Validity (ความถูกต้อง)

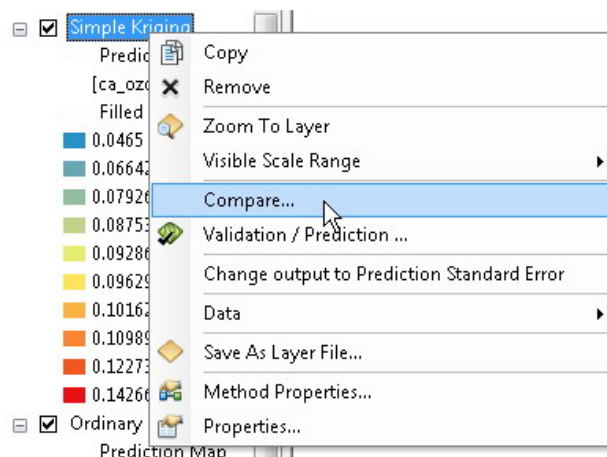
- ต้องตรวจสอบว่า **Root-Mean-Square Standardized** ใกล้เคียงกับ **1** หรือไม่
- เมื่อ **Root-Mean-Square Standardized** ใกล้เคียงกับ 1 และค่า **Average Estimated Prediction Standard Errors** ใกล้เคียงกับ **Root-Mean-Squared Prediction Errors** จากการทำ **Cross-validation** คุณจะมั่นใจได้ว่าโมเดลนั้นเหมาะสมและถูกต้อง

การเลือกโมเดลที่ดีที่สุด

- ในกราฟที่แสดงในตัวอย่าง โมเดล **Kriging** ทางซ้ายมี **Root-Mean-Square** และ **Average Standard Error** ต่ำกว่าของโมเดลทางขวา
- อย่างไรก็ตาม โมเดลทางขวาคงจะได้รับการเลือก เพราะว่า **Root-Mean-Square** และ **Average Standard Error** ใกล้เคียงกัน และ **Root-Mean-Square Standardized** ในโมเดลทางซ้ายมีค่าสูงมาก ซึ่งบ่งชี้ว่าโมเดลนี้มีปัญหาอย่างรุนแรง

ขั้นตอนในการเปรียบเทียบโมเดล

1. คลิกขวาที่ **Geostatistical Layer** ที่คุณต้องการเปรียบเทียบใน **Table of Contents** ของ **ArcMap**
2. เลือก **Compare**
3. เลือก **Layer** ที่สอง ที่จะใช้เปรียบเทียบจากเมนู **To drop-down**
4. คลิกที่ แท็บต่างๆ เพื่อดูผลลัพธ์ที่แตกต่างจากการเปรียบเทียบ
5. คลิก **Close** เมื่อคุณเสร็จสิ้นการเปรียบเทียบ



สรุป

- **Optimality** คือการเลือกโมเดลที่ให้ค่า **Root-Mean-Squared Prediction Error** ต่ำสุด
- **Validity** คือการตรวจสอบว่าโมเดลนั้นเหมาะสมและการคาดการณ์มีความไม่แน่นอนที่สามารถประเมินได้อย่างถูกต้อง
- ควรเปรียบเทียบโมเดลโดยใช้ สถิติ และ กราฟ ที่มีอยู่ใน **Comparison Dialog Box** รวมถึงข้อมูลที่ได้จาก **ESDA** เพื่อการตัดสินใจที่ดีขึ้นในการเลือกโมเดลที่ดีที่สุดสำหรับงานของคุณ!