

STATISTIK 2 (107.325) WS 2010
COMPUTERSTATISTIK (107.258) WS 2010

Übung 10

12. Jänner 2011
Dutter

37. Beschreiben Sie in allgemein verständiger Sprechweise mit eigenen Worten:

- Was bedeutet „Stationarität zweiter Ordnung“?
- Was bedeutet „Intrinsische Hypothese“?
- Was ist der wesentliche Unterschied zwischen den beiden Begriffen?

38. Betrachten Sie die ccnc-Daten, wobei die verschiedenen Variablen nur von der eindimensionalen Variablen „Length“ abhängen. (Die Abkürzung „ccnc“ steht für „Kupfer, Kobalt, Nickel, Chrom“. Werte wurden gemessen entlang einer geraden Linie in Abständen von 50 Meter. Die geografischen Koordinaten werden hier einfach eindimensional mit der Variablen „Length“ ausgedrückt.) Sie stehen auf der Vorlesungshomepage, d.h. man kann sie z.B. mit

```
> addr <- "http://www.statistik.tuwien.ac.at/public/dutt/vorles/geo_data/ccnc.dat"
> ccnc.dat <<- matrix(scan(addr), ncol = 5, byrow = T)
> dimnames(ccnc.dat) <- list(NULL, c("Length", "Cu", "Co", "Ni",
+   "Cr"))
```

einlesen und die Variablen umbenennen.

Wie kann man die Abhängigkeitsstruktur der Variablen im Raum ausdrücken?

Das Variogramm ist unter Stationarität theoretisch definiert als

$$2\gamma(\mathbf{h}) = \text{Var}\{Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})\} = E\{(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}))^2\} .$$

Eine Schätzung dieser „Variogrammfunktion“ wäre in naheliegender Weise durch Ersetzung der Erwartung mit Durchschnitt gegeben:

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{n(\mathbf{h})} \sum_{i=1}^{n(\mathbf{h})} (z(\mathbf{x}_i + \mathbf{h}) - z(\mathbf{x}_i))^2$$

wobei $z(\mathbf{x})$ für den konkreten Datenwert an der Stelle \mathbf{x} steht und $n(\mathbf{h})$ die Anzahl von zur Verfügung stehenden Wertepaaren mit Abstand \mathbf{h} bedeutet.

Nimmt man nun z.B. nur die ersten 15 Werte von Cu, so kann man schon ein Variogramm händisch (oder mit simplen \mathbb{R} -Befehlen) an den Stellen $h = 50m, 100m, \dots, 350m$ schätzen. Bitte, tun Sie das, sodass die Funktion etwa der Abbildung 1 entspricht. Zeichnen Sie auch Ihr Resultat.

39. Betrachten Sie nun alle 102 Messwerte von Cu und ein geeignetes \mathbb{R} -Paket, z.B. `geoR`.

Für die Berechnung mit `geoR` brauchen Sie i.w. die Funktionen `as.geodata` und `variog` und zum Zeichnen ev. `plot.variogram`. Bitte, beachten Sie, dass das Paket

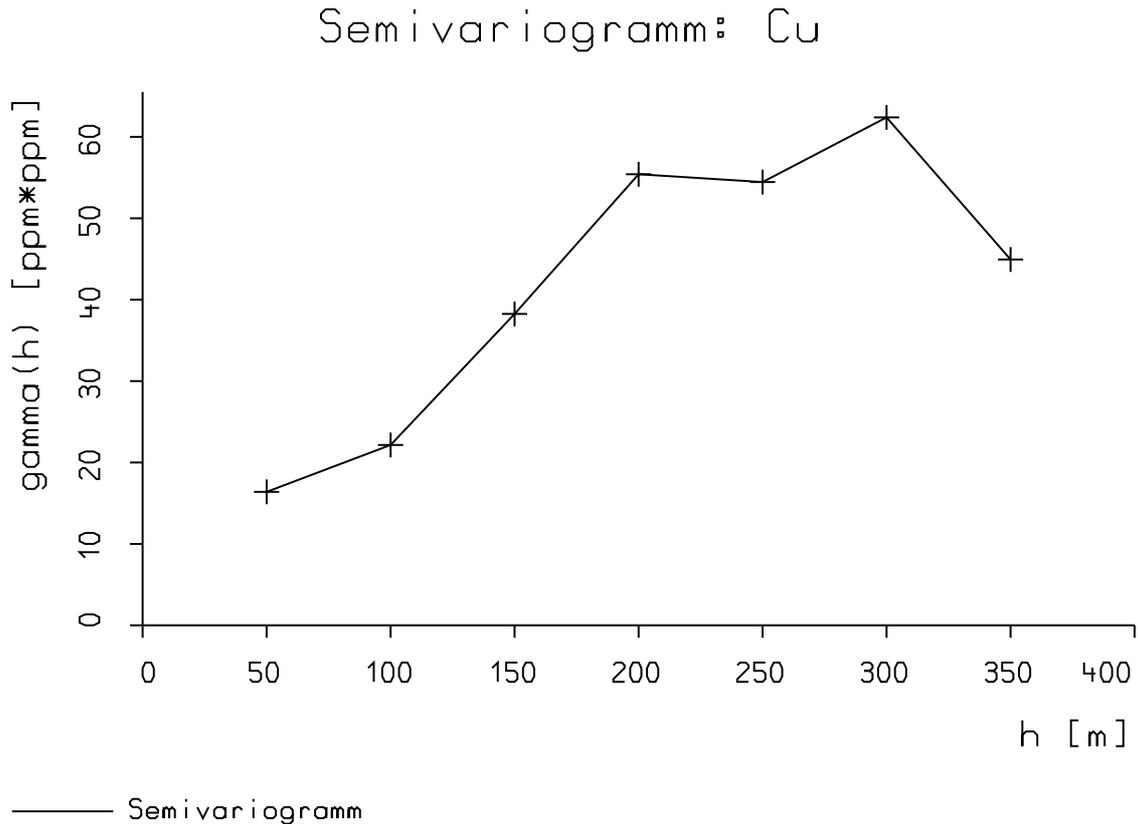


Abbildung 1: Einfaches Variogramm.

nur Daten mit Abhängigkeiten vom 2-dimensionalen geografischen Raum verarbeiten kann. Das heisst, man muss die zur „geografischen Koordinate“ `Length` noch eine künstliche 2. Variable hinzufügen, die z.B. konstant 0 ist. Weiters ist es günstig, für die weitere Rechnung eine neue Datenklasse mit der Funktion `as.geodata` zu konstruieren. Zusammenfassend könnte der Befehl so aussehen:

```
dat <- as.geodata(cbind(ccnc.dat[,Length"],0,ccnc.dat[,Cu]))
```

Die Funktion `variog` zur Schätzung des Variogramms kann dann direkt auf `dat` angewendet werden. Versuchen Sie ein empirisches Variogramm in der Art wie es in Abbildung 2 zu sehen ist, zu erzeugen.

40. Verwenden Sie die Daten „chorizon“, geochemische Daten aus dem C-Horizont des Kolaprojekts, dokumentiert im Paket „mvoutlier“ des Computer Programm Systems  (ähnlich zu den Prüfungsangaben).

(Bem.

```
require("mvoutlier")
data(chorizon)
```

Berechnen Sie ein empirisches Variogramm der Variablen `Fe` in der einfachsten Form (omnidirektional), das etwa so wie in Abbildung 3 aussieht. Diskutieren Sie die Eigenschaften.

Bitte, stellen Sie die Ausarbeitung in Form eines pdf-Files (nicht mehr als 3 Seiten) mit den Resultaten (Outputs plus textliche Kommentare) und Kurz-Listing des Programmkodes (Funktion) in die TUWEL-Seite

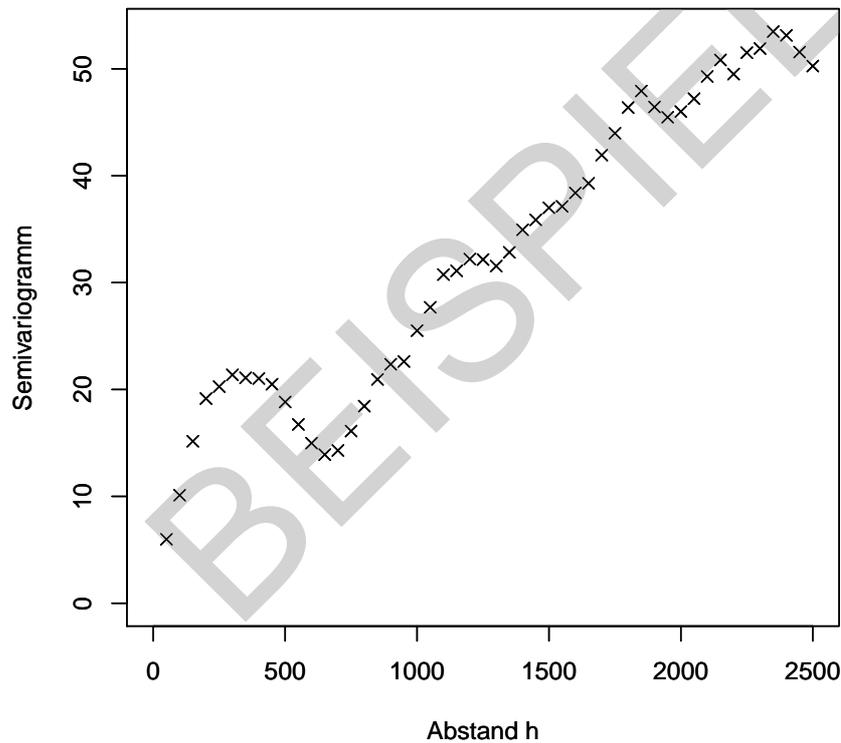
Variogramm der Cu-Daten

Abbildung 2: Variogramm der Cu-Daten von ccnc.

<https://tuwel.tuwien.ac.at/course/view.php?id=2604>
bis incl. 17. Jänner 2011.

Bitte den Namen des pdf-File folgendermaßen:

name_exer_10.pdf

wobei 'name' für den Familiennamen steht. Bitte im pdf-File Name, Datum und Seitennumerierung nicht vergessen!

Empfehlenswert ist es, den Bericht mit „Sweave“ zu erstellen. (Vorlage siehe <http://www.statistik.tuwien.ac.at/public/dutt/vorles/>.)

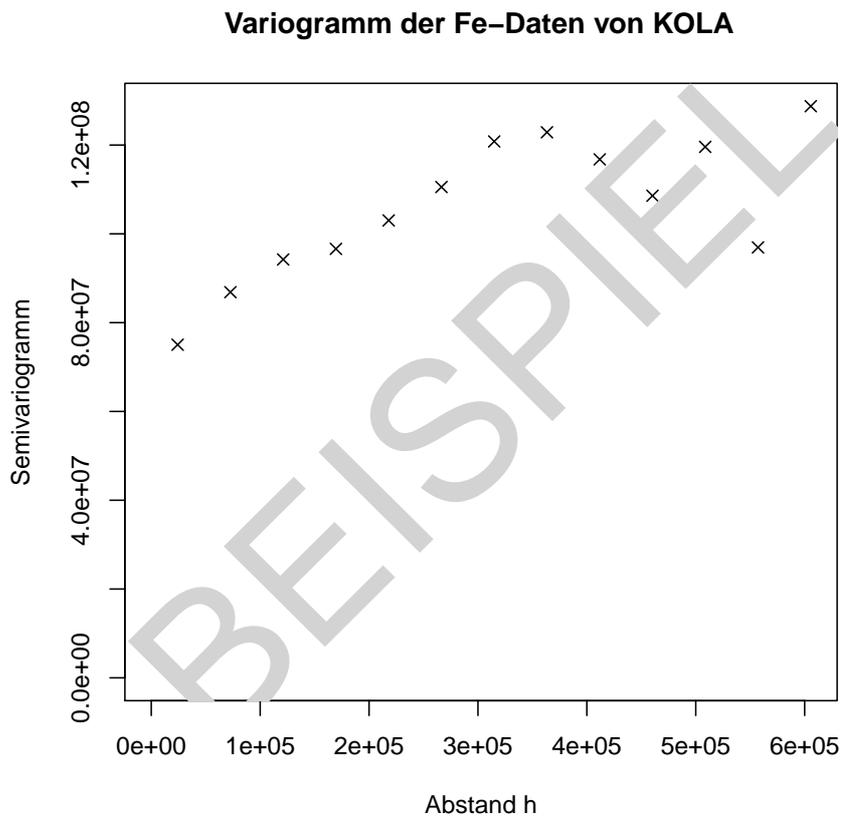


Abbildung 3: Variogramm der Fe-Daten von chorizon des KOLA-Projekts