

## **Progetto di Eccellenza**

### **Attività didattiche di alta qualificazione - A1: Creazione di nuovi insegnamenti**

#### **Relazione del Gruppo di Lavoro (Presidente CCS-LM e referenti d'area)**

Il Progetto di Eccellenza, al Quadro D.7 – punto A1, prevede l'attivazione di 4 nuovi insegnamenti: 1 sul tema dei dati ad elevata dimensionalità, 1 dei dati funzionali, 1 dei dati gerarchici e 1 dei dati da fonti multiple. Il Progetto di Eccellenza prevede l'avvio del processo con la coorte 2019/20.

Stanti gli argomenti di livello avanzato, si tratta di insegnamenti previsti per il secondo anno della LM.

Il 12/7/2018 il coordinatore del Gruppo di Lavoro ha chiesto ai referenti d'area di discutere coi colleghi quali dei 4 insegnamenti fossero di maggiore interesse per l'area, e definirne titolo e SSD. Le aree hanno inoltre proposto dei sillabi sintetici indicativi, che saranno poi rivisti in forma definitiva al momento della loro pubblicazione a cura del docente responsabile. Infatti, trattandosi di insegnamenti che si vanno ad aggiungere all'offerta didattica della LM, così come fu definita e coordinata dalla Commissione Didattica Istruttoria, i programmi proposti non sono vincolanti. Devono essere, in ogni caso, rispettati i temi indicati nel PE e non vi devono essere sovrapposizioni con altri insegnamenti "major" già presenti nell'offerta didattica.

Il gruppo di lavoro si è poi riunito i giorni 18/9/2017, 1/10/2018, 12/10/2018, 19/10/2018 e ha recepito e discusso le indicazioni delle aree in merito ad assegnazioni, titoli e sillabi degli insegnamenti.

Titoli, SSD e brevi sillabi sono stati presentati per la discussione alla Commissione Didattica del 4 ottobre 2018 e al CCS del 9 ottobre 2018 e sono riportati di seguito.

## **Metodi statistici per dati con elevata dimensionalità**

SECS-S/01

1. Verifica di ipotesi su larga scala e test multipli: Bonferroni, FWER e FDR e loro controllo, metodi bayesiani empirici e FDR locale
2. Modelli di regressione per dati con elevata dimensionalità
  - a. dati con numerosità elevata (tall data): algoritmi incrementali a memoria limitata, discesa stocastica del gradiente, inferenza
  - b. dati con un numero elevato di variabili: sparsità, penalizzazioni che inducono sparsità, Lasso, inferenza post-selezione
3. Stima di matrici di varianza e covarianza ad elevata dimensionalità
  - a. modelli fattoriali, modelli che inducono sparsità nella matrice di precisione/concentrazione
  - b. grafi di concentrazione, algoritmi di apprendimento basati su calcoli locali
  - c. applicazioni in ambito finanziario e bioinformatico

## **Metodi statistici per dati funzionali**

SECS-S/01

1. Metodi esplorativi
2. Modelli frequentisti - lisciamento di funzioni e funzioni vincolate - registrazione di curve - metodi di riduzione della dimensionalità (componenti principali di funzioni e relative regolarizzazioni) - modelli con esplicative funzionali - modelli per risposta funzionale ed esplicative multivariate - modelli per risposta funzionale ed esplicative funzionali e modelli concorrenti - modelli per equazioni differenziali
3. Modelli bayesiani - modelli parametrici - modelli non parametrici (BNP)

## **Metodi statistici per l'analisi di dati gerarchici**

SECS-S/03

Il corso si propone di introdurre metodologie e tecniche avanzate per il trattamento e l'analisi di dati aventi una struttura gerarchica.

Dopo alcuni richiami sui principali concetti e strumenti per lo studio della variabilità intra e tra gruppi e per l'analisi di dati gerarchicamente strutturati (in particolare, i modelli multilivello), il corso approfondirà da una parte temi aventi una forte connotazione pratica (caratterizzazione della struttura gerarchica in funzione dell'obiettivo dell'analisi, il ruolo della centratura delle variabili in un contesto multilivello e gli effetti delle numerosità campionarie dei diversi livelli sull'accuratezza delle stime). Dall'altra verrà introdotta l'analisi multilivello in una struttura generale a variabili latenti.

Numerose applicazioni in ambito economico, finanziario e del marketing aiuteranno la comprensione e l'approfondimento di questi temi.

## **Dati multi-fonte e analisi territoriali**

SECS-S/05

Sempre più spesso, le basi di dati statistiche nascono dall'integrazione di archivi diversi, costruiti con dati provenienti da fonti sia di tipo "tradizionale" (censimenti, dati amministrativi, dati da indagini campionarie) che "moderno" (web, social network, droni-telecamere).

È quindi opportuno che gli studenti della laurea magistrale di scienze statistiche – anche in vista del loro impegno lavorativo – acquisiscano le competenze necessarie per costruire le basi di dati multi-fonte.

La prima parte di questo corso sarà dedicata alle procedure di linkage, di gestione della selezione, di identificazione e correzione degli errori, di gestione degli outliers. Tali procedure statistiche (deterministiche e probabilistiche) verranno apprese attraverso un'intensa attività laboratoriale, grazie all'analisi di basi di dati reali.

Gran parte delle basi di dati multi-fonte interessa singoli o aggregati dislocati in un territorio. La seconda parte del corso sarà dedicata alla rappresentazione dei dati territoriali, sia mediante l'acquisizione delle tecnologie GIS, sia attraverso tecniche di cartografia statisticamente determinate. Anche questa parte verrà sviluppata mediante un'intensa attività laboratoriale.