

Temas avanzados en física computacional: Análisis de Datos

Proyecto Final Semestre académico 2016-I

Elaborado por Prof. J. Bazo

Tema general del proyecto.-

Proceso completo del análisis de datos aplicado a un caso real

Duración del proyecto.-

El proyecto durará 6 semanas terminando el 1° de julio con una presentación. La mayor parte del trabajo se desarrollará fuera del horario de clases.

Objetivos del proyecto.-

- Poner en práctica las herramientas computacionales y estadísticas aprendidas durante el curso en un caso específico y que por su complejidad sea relevante.
- Desarrollar un código reproducible.
- Incentivar y desarrollar el trabajo en equipo que es indispensable en la generación de grandes proyectos de código y análisis de datos.

Preguntas guía.-

- ¿Queremos utilizar datos públicos, mediciones propias o simulaciones?
- ¿Cuál ambiente de programación manejamos mejor, R o ROOT?
- ¿Qué set de datos se presta mejor para aplicar un análisis multivariado?

Etapas del proyecto.-

1. Planeamiento: selección del tema y alternativas de organización
2. Recolección y procesamiento de datos
3. Análisis estadístico y multivariado
4. Preparación de ayuda visual para la presentación
5. Reporte final oral

Instrucciones.-

1. Seleccionen un tema y una pregunta específica donde se pueda aplicar el análisis de datos y en especial un análisis multivariado, así como la fuente de los datos a utilizar y el ambiente de programación. Distribuyan las tareas entre los miembros del grupo y definan las fechas límites para cada actividad. Como evidencia de esta parte elaborar un pequeño informe inicial sobre los puntos de la planificación que deberá ser entregado a más tardar el 3 de junio.
2. En las tres siguientes actividades utilicen github o un sistema equivalente para el seguimiento de las versiones de sus códigos y poder compartir éstos entre los miembros del grupo y el profesor. Además, cada vez que el código de un participante esté en su etapa final, otro miembro del grupo lo deberá revisar (code review) y subir una nueva versión.
3. Recolecten y procesen los datos de acuerdo a la pregunta planteada, según el rol de los integrantes del grupo y la fecha límite prevista. Dependiendo del caso conservar un script único que realice toda esta etapa e incluya una descripción de las variables, unidades, cortes, fecha de los datos, etc.
4. Decidan las pruebas estadísticas a realizar y el tipo de análisis multivariado que usarán. Desarrollen el código necesario y si el método no funciona para el caso específico, busquen alternativas. Si la división de tareas iniciales y tiempo asignado requiere un cambio este será posible, pero hay que comunicarlo o escribirlo en la autoevaluación.
5. Preparen gráficos pertinentes que se expliquen individualmente y que representen los pasos intermedios y resultados finales del proyecto.
6. Preparen una presentación final que incluya los diversos pasos del proyecto, las alternativas de solución y también aquellas descartadas o que no tuvieron un resultado positivo, los resultados obtenidos y conclusiones. En la presentación oral, que deberá durar entre 30 minutos y 1 hora, deberán participar todos los integrantes.
7. Escriban una autoevaluación sobre la participación de los integrantes, el cumplimiento de las fechas propuestas, las fortalezas y debilidades del trabajo en equipo. Esta autoevaluación se entregará el mismo día de la presentación.

Criterios de evaluación del proyecto.-

1. Informe inicial sobre planificación y elección del tema (2ptos)
Se evalúa la organización del grupo (selección de la pregunta, relevancia actual de ésta, reparto eficiente de tareas y delimitación de tiempos en el informe inicial.
2. Recolección y procesamiento de datos (3ptos)
Se evalúa la habilidad para recolectar y procesar datos (archivo de datos procesados estructurado, descripción clara de las variables, código comentado)

3. Manejo de herramientas estadísticas y de análisis multivariado (5ptos)
Se evalúa el número de métodos utilizados, la complejidad de los parámetros, el código comentado, las funciones/distribuciones adicionales usadas no explicadas en clase.
4. Calidad de los gráficos obtenidos (3ptos)
Se evalúan los ejes con títulos claros, los colores adecuados, la claridad del contenido gráfico, las leyendas, el uso de diversos formatos y la creatividad.
5. Presentación final (4ptos)
Se evalúa la participación de todos los integrantes, la organización de la presentación, la claridad, las alternativas de solución descartadas o que no tuvieron un resultado positivo, los resultados obtenidos y las conclusiones.
6. Proceso de trabajo en equipo y autoevaluación (3ptos)
Se evalúa la colaboración en el desarrollo del código vía github (código almacenado paulatinamente), la revisión del código de un participante por otro participante y la autoevaluación (honestidad e identificación de fortalezas y debilidades).

San Miguel, 20 de mayo de 2016.