**Instituto de Nutrición de Centro América y Panamá**

**Centro de Investigación del INCAP para la Prevención de Enfermedades Crónicas**

**Curso presencial: Introducción a la Programación Estadística**

**Duración: 17 de agosto 2017 al 23 de noviembre 2017**

|  |
| --- |
| **Al finalizar el participante tendrá la capacidad de:**  Proveer soporte estadístico computacional utilizando software de licencia gratuita aplicado en investigación epidemiológica y nutricional. |
| **Evidencia de aprendizaje:**  Los archivos de Jupyter Notebook que incluyen el código programado según los casos presentados. |
| **Indicador de logro:**  Análisis de base de datos de prueba desde limpieza de datos hasta creación de mapas en busca de relaciones de variables. |

**Dr. Kevin Martinez-Folgar**

**kmartinez@incap.int**

|  |  |  |  |
| --- | --- | --- | --- |
| **Unidad I: Introducción a Python** | | | |
| **Tema** | **Objetivos** | **Recursos** | **Actividades de enseñanza-aprendizaje** |
| 1. Definición de Python 2. ¿Por qué aprender Python para la ciencia? 3. Instalación de Python en su paquete científico 4. Introducción a Jupyter/IPython 5. Jupyter Notebook 6. Introducción a la Programación de Objetos 7. Sintaxis de Python básica | **Cognoscitivo:**  Comprender la importancia del uso de software con licencia libre para el análisis de datos.  **Procedimental:**  Utilizar la interfaz de Jupyter notebook para programación en Python.  Aplicar sintaxis básica de Python para realizar operaciones aritméticas dentro de Jupyter. | **Conferencia inaugural**  “Introducción a Python para el análisis de datos”  **Fecha**: 17 agosto 2017  **Hora:** 10:00 hrs (Guatemala)  **Lecturas complementarias**   1. McKinney W. (2012) ***Python for Data Analysis***. Capitulo 1. Editorial O’Reilly. 2. McKinney W. (2012) ***Python for Data Analysis***. Appendix: Python language Essentials. Editorial O’Reilly. | Copia de los archivos de Jupyter Notebook compartidos. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Unidad 2: Introducción a Numpy y Pandas** | | | |
| **Tema** | **Objetivos** | **Recursos** | **Actividades de enseñanza-aprendizaje** |
| 1. Introducción a NumPy    1. Numpy Arrays    2. Arrays de Dos dimensiones    3. Valores “missing” 2. Introducción a Pandas    1. Objeto “Series” de Pandas    2. Objeto “DataFrame” de Pandas 3. Importación de bases de datos    1. CSV    2. STATA    3. Excel 4. Exploración de la base de datos y aplicación de comandos básicos    1. info    2. describe    3. shape    4. columns    5. head    6. tail    7. sum    8. count\_values 5. Índices y Mascaras 6. Valores “missing” | **Cognoscitivo:**  Comprender la librería de NumPy y Pandas utilizando arrays, series y dataframes aplicando ejercicios prácticos  **Procedimental:**  Analizar los casos de interés de la base de datos utilizando índices y máscaras.  Utilizar la interfaz de Jupyter notebook para programación en Python. | **Conferencia y Taller**  “Introducción a NumPy y Pandas para el análisis de datos”.  **Fecha**: 31 agosto 2017  **Hora:** 10:00 hrs (Guatemala)  **Lecturas complementarias**   1. McKinney W. (2012) ***Python for Data Analysis***. Capitulo 4 O’Reilly Media. | Copia de los archivos de Jupyter Notebook |
| **Unidad 3: Manipulación de datos con Pandas** | | | |
| **Tema** | **Objetivos** | **Recursos** | **Actividades de enseñanza-aprendizaje** |
| 1. Columnas (Variables)    1. Crear columnas    2. Renombrar columnas    3. Eliminar columnas    4. Explorar columnas de texto (string) 2. Filas (Observaciones)    1. Eliminar observaciones    2. Etiquetar variables categóricas    3. Codificar variables categóricas 3. Agrupar variables (Group By) 4. Tablas de Variables Cruzadas 5. Segmentar bases de datos 6. Unir bases de datos (Merge/Join) | **Cognoscitivo:**  Explicar el manejo de datos con Pandas.  **Procedimental:**  Manipular la base de datos utilizando tablas de variables cruzadas y variables agrupadas. | **Conferencia y Taller**  “Manipulación de datos con Pandas”.  **Fecha**: 13 septiembre 2017  **Hora:** 10:00 hrs (Guatemala)  **Lecturas complementarias**   1. McKinney W. (2012) ***Python for Data Analysis***. Capitulo 7 O’Reilly Media. | Copia de los archivos de Jupyter Notebook |

|  |  |  |  |
| --- | --- | --- | --- |
| **Unidad 4: Visualización de datos** | | | |
| **Tema** | **Objetivos** | **Recursos** | **Actividades de enseñanza-aprendizaje** |
| 1. Graficando con Pandas 2. Grafica de barras 3. Histograma 4. Gráficos de caja 5. Introducción a matplotlib 6. Grafico Linear simple 7. Gráfico de dispersión simple 8. Curvas de densidad. 9. Graficando con Seaborn 10. Graficas utilizando archivos Shapefiles (Mapas)     1. GeoPandas | **Cognoscitivo:**  Comprender la librería de matplotlib y seaborn para la creación de gráficas.  **Procedimental:**  Producir graficas utilizando Pandas, matplotlib, seaborn.  Construir mapas de colores *(choroplet map)* utilizando la base de datos. | **Conferencia y Taller**  “Visualización de datos”  **Fecha**: 28 septiembre 2017  **Hora:** 10:00 hrs (Guatemala)  **Lecturas complementarias**   1. McKinney W. (2012) ***Python for Data Analysis***. Capitulo 8 O’Reilly Media. | Copia de los archivos de Jupyter Notebook compartidos. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Unidad 5: Estadística Descriptiva** | | | |
| **Tema** | **Objetivos** | **Recursos** | **Actividades de enseñanza-aprendizaje** |
| 1. Medidas de tendencia central    1. Media    2. Mediana    3. Moda 2. Medidas de posición    1. Cuartiles    2. Quintiles    3. Deciles    4. Percentiles    5. “Outliers” 3. Medidas de dispersión    1. Desviación estándar    2. Varianza | **Cognoscitivo**   * Describe distribución de los datos utilizando medidas de tendencia central, posición y dispersión. * Identifica valores aberrantes dentro de la distribución.   **Procedimental.**   * Utiliza Python para realizar cálculos estadísticos descriptivos. | **Conferencia y Taller**  “Estadística descriptiva”  **Fecha**: 12 octubre 2017  **Hora:** 10:00 hrs (Guatemala)  **Lecturas complementarias**   1. McKinney W. (2012) ***Python for Data Analysis***. Capitulo 9 O’Reilly Media. | Copia de los archivos de Jupyter Notebook compartidos. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Unidad 6: Estadística Inferencial** | | | |
| **Tema** | **Objetivos** | **Recursos** | **Actividades de enseñanza-aprendizaje** |
| 1. Test de Hipótesis    1. Identificación de representatividad de nuestra base de datos. 2. Relaciones entre variables    1. Gráficos de dispersión    2. Correlación    3. Covarianza    4. Correlación de Pearson    5. Relaciones no lineares    6. Correlación de Spearman 3. Diferencia entre grupos    1. Diferencia de medias    2. Test de Proporciones    3. Chi-Cuadrado 4. Estimación    1. Error Estándar    2. Intervalos de confianza 5. Tamaño del Efecto    1. D de Cohen | **Cognoscitivo**   * Comparar la relación entre variables y la comprobación de hipótesis estadísticas. * Ejemplificar pruebas estadísticas aplicables para comprobar una hipótesis.   **Procedimental**   * Calcular las pruebas estadísticas para resolver las pruebas de hipótesis. | **Conferencia y Taller**  “Estadística inferencial”  **Fecha**: 26 octubre 2017  **Hora:** 10:00 hrs (Guatemala) | Copia de los archivos de Jupyter Notebook compartidos. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Unidad 7: Regresión** | | | | | |
| **Tema** | **Objetivos** | | **Recursos** | | **Actividades de enseñanza-aprendizaje** |
| 1. Predicción de relación entre variables    1. Regresión Linear Simple    2. Regresión Múltiple    3. Regresión Logística | **Cognoscitivo**   * Identificar la distribución de los datos para seleccionar método inferencial apropiado. * Interpretar correctamente los resultados de una regresión.   **Procedimental**   * Computar comandos para generar los diferentes tipos de regresión. | | **Conferencia y Taller**  “Regresión”  **Fecha**: 9 noviembre 2017  **Hora:** 10:00 hrs (Guatemala) | | Copia de los archivos Jupyter Notebook compartidos. |
| **Unidad 8: Ruta de trabajo para el análisis de bases de datos** | | | | | |
| **Tema** | | **Objetivos** | | **Recursos** | **Actividades de enseñanza-aprendizaje** |
| 1. Secuencias para el análisis de bases de datos 2. Importar y limpiar la base de datos 3. Explorar una variable 4. Explorar variables utilizando tabulación de resultados 5. Realizar un análisis multivariado 6. Estimar el test de hipótesis y parámetros asociados 7. Visualización de resultados utilizando métodos gráficos | | **Cognoscitivo**   * Distinguir los pasos necesarios para analizar una base de datos.   **Procedimental**   * Practicar los pasos necesarios para analizar una base de datos. | | **Conferencia y Taller**  “Análisis de Bases de Datos”  **Fecha**: 23 noviembre 2017  **Hora:** 10:00 hrs (Guatemala) | Copia de los archivos Jupyter Notebook compartidos. |