Actividad:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Unidad** | **Módulos** | **Temas** | **Objetivos** | **Screencasts** | **Observaciones** |
| **U1. Preparación del Ambiente de Trabajo e Introducción de Conceptos Fundamentales de R, RStudio y RMarkdown**  (3 horas) | a. Instalación de de RStudio con R, paquetes relacionados y dejar listo para trabajar. | 1. Prepara el ambiente en PC (incluye Surface) |  | Instalación en PC | Completado |
|  |  | 2. Preparar el ambiente en Mac |  | Instalación en Mac | Completado |
|  |  | 3. Preparar el ambiente en Linux |  | Instalación en Linux | Más tarde. |
|  | b. Familiarización con el interfaz de RStudio | 1. Explicación general del interfaz de RStudio. |  | a. **Ventana de Consola**: Ejemplo básico de hacer aritmética en la consola y explicar propósito.  b. **Ventana de Script**: Crear y editar un *Script* para ejecutar operaciones más complejas y explicar propósito y ventajas.  c. **Ventana de Archivos y Paquetes**: Utilidad de los paquetes, manejo de archivos y explicar propósito.  d. **Ventana del Ambiente**: Explicación general de variables, explicación de visualización de otros objetos, importación de datos. |  |
|  |  | 2. Propósito de cada ventana. |  |  |  |
|  |  | 3. Utilidad de cada ventana. |  |  |  |
|  | c. Configuración de directorio y manejo de archivo | 1. Creación de carpeta en unidad propia. |  | Creación de carpeta en Mac, PC y Linux. |  |
|  |  | 2. Configuración de RStudio para utilizar un directorio de trabajo |  | Configuración con Mac, PC, Linux |  |
|  |  | 3. Incorporación de un archivo al directorio de trabajo | Dado un archivo enviado previamente, incorporará el archivo al directorio de trabajo y lo verificará en RStudio. | Incorporación del archivo de Finder (Mac), Explorer (PC), (Linux) y comprobación de la existencia del archivo en el directorio de trabajo.  Demostrar cómo localizar el archivo de datos en PC, Mac, Linux |  |
|  | d. Instalación de paquetes | 1. Utilizar el diálogo “Install Packages” para instalar paquetes al ambiente | Dada una lista de paquetes previamente entregada, los instalarán utilizando la Ventana de Archivos y Paquetes. | Instalación de una serie de paquetes del depósito CRAN: ggplot2, <lista de Denny>. Encontrar paquete en **Ventana de Archivos y Paquetes** y corroborar que se instaló. |  |
|  |  | 2. Visualización del paquete |  |  |  |
|  | e. Introducción a RMarkdown | 1. Crear un archivo en RMarkdown y guardar en directorio de trabajo. | Crearán y editarán un archivo en RMarkdown, añadirán una imagen, visualizarán el resultado.  Abrir un archivo en RMarkdown | Identificar cómo luce un documento en Rmarkdown de acuerdo a sus características en el encabezado del documento.  Demostración de cómo escribir texto, incorporar texto, escribir comentarios, incorporar imagen y visualizarla en HTML y MS-Word. |  |
|  |  |  |  | Demostrar cómo abrir un documento en RMarkdown |  |
|  |  | 2. Creación y manejo de un *chunk* de R. | Crear un chunk, identificar sus características, documentar con comentarios lo que hace. Correr chunk y visualizar resultados. | Demostrar cómo:   * crear un chunk, * identificar sus partes, * ejecutarlo (por línea, por pedazo, todo a la vez, etc.) y * visualizar el resultado. * Demostrar cómo documentar (poner comentarios) dentro de un chunk. | Considerar ejercicio demostrando cómo calcular el área de un círculo utilizando fórmulas matemáticas en R. |
|  |  | 3. Edición básica de texto, imágenes, etc. en RMarkdown | Manipular texto en RMarkdown con los códigos requeridos. | Demostrar cómo un texto dado se le ponen negrillas, itálicas y subrayado. Incorporar imágenes.  Crear enlaces a un URL. | Pendiente |
|  |  | 4. Visualización de documentos creados en RMarkdown. | Manipular la visualización en la creación del archivo de salida. | Mostrar cómo visualizar (tejer o “knit”) el resultado en HTML, Word, y pdf.  Hacer modificaciones a lo anterior. | NOTA: Posibilidad de crear muchos ejemplos de acuerdo al tema y la audiencia. |
|  | f. Introducción de datos en el ambiente de R | 1. Incorporación de datos al ambiente de trabajo:  \* objetos  \* tipos de datos (variables, etc.)  \* vectores  \* data.frame  \* matrices  \* otros tipos de estructuras | Incorporará datos manualmente. Los importará.  Los visualizará.  Podrá nombrar y diferenciar diferentes tipos de estructuras de datos en R. | Demostrar cómo:   * Introducir datos manualmente a un vector * Crear una matriz a partir de un vector * Crear un data.frame a partir de vectores * Visualizar los datos en la **Ventana del Ambiente** y mediante la **Consola**. | Puede que requiera varios segmentos de vídeo de no más de cinco minutos de duración. |
|  |  | 2. Importación de datos al ambiente de trabajo: | Importará datos de hojas de cálculo, definirá parámetros de importación, utilizará la ventana “Code Preview” | Demostrar importación de archivos   * manualmente directamente de hojas de cálculo (Excel, Spreadsheet, Numbers, etc.), explicar cómo preconfigurar el archivo antes de incorporar.  Definir parámetros como doubleprecision, caracter, subhoja (como *sheet*, etc.) * usando el código de la ventana de “Code Preview” | Debe salir en cinco minutos. Se pueden crear otros con diferentes variaciones. Unos screencasts adicionales con más detalle con diferentes problemas demostrados. |
|  |  | Importará datos a través de un API. | Demostrar incorporación de archivos utilizando un AP. | Ej. Importar datos directo de Google Drive. |
|  |  | Importará archivos de otros programas estadísticos (SPSS. SAS, Minitab, etc.) | Demostrar incorporación de archivos de otras aplicaciones (SPSS, SAS, Minitab, etc.) | Conseguir archivo prestado y demostrar cómo traérlo al ambiente. Utilizar un archivo del sitio de Internet de los mismos programas y hacer la demostración. |
| **U2. Estadísticas Descriptivas en R (3 horas)** | a. Distribución de datos | 1. Creación de histogramas y visualización | Creará un histograma y cambiará algunos parámetros con varios datasets. | Importar datos de un archivo dado, escoger variable, crear histograma con ggplot2, función geom\_histogram. Manipular algunos parámetros básicos: color, curva de densidad, nombre a los ejes, añadir geom\_rug |  |
|  |  |  |  | Demostrar cómo entrar los datos manualmente a un data.frame (no importándolos directamente de un archivo CSV, Excel, etc.) | Pendiente a ubicar en alguna unidad. |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| **U3. Pruebas Estadísticas** | Pruebas estadísticas básicas | Regresión lineal simple | Ejecutar una prueba de regresión lineal simple. | * Demostrar una prueba de regresión lineal simple * Interpretación de prueba de regresión lineal simple | Un screencast para demostrar cómo montar la prueba, otro para la interpretación. |
| Probar supuestos de la regresión lineal simple. | * Demostrar cómo probar los supuestos.   + Normalidad con *Q-Q Plot,* visualizar e interpretar los resultados.   + Prueba *Shapiro-Wilk*, visualizar e interpretar los resultados. | Un solo screencast. Si no sale en cinco minutos, hacer uno por cada prueba. |
| Mejorar el ajuste de acuerdo al resultado de la R2 | * Demostrar función polinomial   + Con el elemento cuadrático   + Reinterpreta los resultados de la R2 | En cinco minutos. |
| ANOVA de una vía - básica |  | Dos screencasts para los siguiente:   * Introducción básica de la ANOVA de una vía. * Demostración con unos datos dados. * Correr prueba, interpretar resultados. * Correr posthoc | Demostrar cómo correr la prueba e interpretarla.  Correr pruebas de supuestos  Correr análisis posthoc |
|  | Pruebas estadísticas intermedias |  | Cómo hacer una prueba de normalidad. | * Demostrar cómo instalar y activar el paquete *EnvStats* (instalar, corroborar que funciona, etc.). * Correr la función *Q-Q Plot* para visualizar gráficamente la normalidad. Interpretar los resultados. * Correr la prueba *Shapiro-Wilk*, visualizar e interpretar los resultados. | Ver U1.d.1. |
|  |  |  | Cómo hacer una prueba de linealidad. | * Demostrar cómo instalar y activar paquete *car* * Correr la prueba *boxTidwell* de linealidad. Interpretar los resultados |  |
|  |  |  | Cómo hacer una prueba de autocorrelación | * Comprobar que paquete *car* esté instalado y corriendo. * Correr la prueba *Durbin-Watson* para autorcorrelación. Visualizar e interpretar los resultados. |  |
|  |  |  | Cómo hacer un análisis de multicolinealidad. | * Comprobar que paquete *car* esté instalado y corriendo. * Correr la prueba *cor* de correlación. Visualizar e interpretar los resultados. * Demostrar análisis *scatterplotMatrix*. Visualizar e interpretar los resultados. |  |
|  |  |  | Utilizar RPubs | Demostrar cómo recuperar y publicar documentos en RPubs. |  |
| **Unidad** | **Módulos** | **Temas** | **Objetivos** | **Screencasts** | **Observaciones** |
|  |  | Modelos de regresión lineal múltiple | Demostrar el modelo de regresión múltiple con modelos, funciones y parámetros específicos. | * Correr la función *modRM*. Correr el comando *summary*. Interpretar los resultados. * Correr la prueba *ncvTest* de homocedasticidad.  Interpretar los resultados. * Visualizar los resultados gráficamente con el comando *spreadLevelPlot*. Interpretar los resultados. * Correr el mismo análisis eliminando variables que demuestren menos correlación.  Interpretar los resultados. * Correr el mismo análisis mostrando la interacción entre variables con el símbolo de dos puntos ( : ). Interpretar los resultados.  Seleccionar el mejor modelo. * Demostración del modelo *stepwise*.  Correr la función *step* que quite y añada variables automáticamente.  Interpretar los resultados. Seleccionar el mejor modelo. * Demostrar comparación de todos los modelos.  Comprobar instalación de paquetes *coefplot* y *ggplot2*.  Correr el comando *multiplot* en varios modelos. Visualizar e interpretar los resultados.  Seleccionar el mejor modelo. * Demostrar la visualización de los R-cuadrado ajustados. Comprobar la instalación del paquete *leaps*. Correr el comando *regsubsets*. Visualizar e interpretar los resultados.  Seleccionar el mejor modelo. * Demostrar el parámetro *Mallow Cp*. Visualizar e interpretar los resultados. Seleccionar el mejor modelo. | No sé a qué unidad/módulo acomodarlo.  Requiere datos que tienen que ser distribuidos de antemano. |
| **Unidad** | **Módulos** | **Temas** | **Objetivos** | **Screencasts** | **Observaciones** |
|  |  | ANOVA de una vía - intermedia | * Diseñar una estrategia para preparar los datos para hacer una prueba de ANOVA de una vía. * Llevar a cabo una prueba del análisis de acuerdo a los datos. * Interpretar resultados de ANOVA de acuerdo a cómo se producen en R. * Hacer análisis de normalidad, homogeneidad y si los residuales tienen una distribución normal. Explicar su importancia en la interpretación de los resultados de la ANOVA. * Hacer un análisis post-hoc para determinar la diferencia entre los grupos. |  Explicar el fundamento estadístico de la ANOVA de una vía, el modelo matemático correspondiente, fórmulas correspondientes, supuestos correspondientes, etc. (puede que esta teoría no se deba incluir en este mismo módulo).   Demostrar cómo entrar los datos manualmente a un data.frame (no leyéndolos directamente de un archivo CVS, Excel, etc.) de manera que facilite una prueba de ANOVA. (hacer referencia a un screencast creado arriba)   * Creación de vectores (hacer referencia a un scrrencast creado arriba) * Creación del data.frame con los vectores (hacer referencia a un scrrencast creado arriba)    Demostrar cómo visualizar los datos de los grupos con un diagrama de caja (*boxplot*) con *ggplot2* con el comando *geom\_boxplot* y hacer la interpretación correspondiente. (Hacer referencia a otro módulo que explica el comando *geom\_boxplot*.)   Correr el análisis de ANOVA con el comando *aov*. Producir un resumen de la prueba con *summary*. Interpretar los resultados.   Preparar los datos para analizar los residuales para mostrar normalidad.  Demostrar cómo hacer prueba Q-Q plot con *qqPlot* e interpretar el gráfico de Q-Q.   Correr la prueba de normalidad Shapiro-Wilk con comando *shapiro.test* para determinar si los datos tienen una distribución normal. Interpretar los resultados.   Correr la prueba *bartlett.test* para demostrar la homogeneidad de los datos. Interpretar los resultados.   Hacer un análisis visual con intervalos de confianza para determinar la diferencia entre las medias de los grupos utilizando *ddply*, *plyr* y *ggplot2*. Interpretar los resultados.   Hacer un análisis post-hoc de Tukey con el comando *TukeyHSD*.  Producir los intervalos de confianza con *plot.* Interpretar los resultados.   Hacer un análisis post-hoc de Bonferroni. Comprar los resultados.   Hacer un análisis post-hoc de Scheffé. Comprar los resultados. | Hacer los siguientes screencasts:   * Introducción explicando diferentes tipos de ANOVAs   + Con repeticiones   + Dos vías   + ANCOVA   + MANOVA * Demostrar cómo preparar los datos para una ANOVA,  rearreglar los datos con *dplyr* * Corre la prueba, interpretar los resultados. |
|  |  | ANOVA con repeticiones | * Diseñar una estrategia para preparar los datos para hacer una prueba de ANOVA de una vía. * Llevar a cabo una prueba de ANOVA de repeticiones de acuerdo a los datos. * Interpretar resultados del análisis de acuerdo a cómo se producen en R. | * Explicar el fundamento estadístico de la ANOVA con repeticiones * Demostrar cómo visualizar los datos de los grupos con un diagrama de caja (“boxplot”) con el comando *boxplot* y hacer la interpretación correspondiente. (Hacer referencia a otro módulo que explica el comando *boxplot*.) * Correr el análisis de ANOVA con el comando *aov*. Producir un resumen de la prueba con *summary*. Interpretar los resultados. * Correr la prueba de normalidad Shapiro-Wilk con comando *shapiro.test* para determinar si los datos tienen una distribución normal. Interpretar los resultados. * Correr la prueba *bartlett.test* para demostrar la homogeneidad de los datos. Interpretar los resultados. | Se podría combinar con la ANOVA de una vía para no repetir los mismo. |
| **Unidad** | **Módulos** | **Temas** | **Objetivos** | **Screencasts** | **Observaciones** |
|  |  | ANOVA con dos factores | * Llevar a cabo una prueba de ANOVA con dos factores de acuerdo a los datos. * Interpretar resultados del análisis de acuerdo a cómo se producen en R. | * Explicar el fundamento estadístico de la ANOVA con dos factores. * Demostrar cómo organizar los datos para llevar a cabo la prueba. * Demostrar cómo verificar (visualizar) los datos crudos con la función *table*. * Demostrar la función *aggregate* para preparar los datos antes del análisis. Visualizar los datos. * Demostrar la función *aggregate* que muestre la media, desviación estándar, etc. Visualizar los datos en cada caso. * Correr el análisis de ANOVA con el comando *aov*. Producir una visualización de las medias con el comando *plotmeans* junto con el intervalo de confianza. Interpretar los resultados. |  |
|  |  | ANCOVA con un factor y una covariable | * Llevar a cabo una prueba de ACOVA con un factor y una covariable de acuerdo a los datos. * Interpretar resultados del análisis de acuerdo a cómo se producen en R. | * Explicar el fundamento estadístico de la ACOVA con un factor y una covariable. * Importar la librería *multcomp* (si no se ha explicado antes). * Demostrar cómo verificar (visualizar) los datos crudos con la función *table*. * Demostrar la función *aggregate* para preparar los datos antes del análisis. Visualizar los datos. * Demostrar la función *aggregate* que muestre la media. Visualizar los datos. * Correr el análisis de ACOVA con el comando *aov*. Producir una visualización del resultado con *summary*. Interpretar los resultados. * Demostrar cómo recalcular (reajustar) las medias al remover la covariable para obtener las medias sin el pseudofactor. |  |
|  |  | MANOVA de un factor y varias niveles | * Llevar a cabo una prueba de MANOVA con un factor y varios niveles de acuerdo a los datos. * Interpretar resultados del análisis de acuerdo a cómo se producen en R. |  Visualizar los datos.   Usar función *cbind* para identificar las variables dependientes.   Demostrar la función *aggregate* que muestre la media. Visualizar los datos.   Correr el análisis de MANOVA con el comando *manova*. Producir una visualización del resultado con *summary*. Interpretar los resultados.   Demostrar cómo identificar el efecto específico de cada variable que causa la varianza.  Interpretar los resultados. |  |
|  |  |  | Configurar RStudio |  |  |
|  |  |  | Configurar RMarkdown |  |  |
| **Unidad** | **Módulos** | **Temas** | **Objetivos** | **Screencasts** | **Observaciones** |
| ¿? | ¿? | Variables categóricas (binaria o politoma) | Explicar sus características, propiedades, lo que describen. etc. | ¿? | ¿? |
| ¿? | ¿? | Regresión Logística | Explicar el uso de la función *glm* de R | * Demostrar una gráfica con una curva logística con *ggplot*. Explicar resultados. * Contrastar variables continuas y variables categóricas dentro de un mismo análisis. * Uso de la función glm para probar modelos de regresión logística. Utilizar el parámetro *family=bionomial=logit*. Explicar la interpretación de los resultados * Obtener los coeficientes.  Explicar la interpretación de los resultados. * Demostrar proceso de aprendizaje automático (“machine learning”).   + Crear un modelo usando la función *glm* con datos de aprendizaje.  Demostrar el nivel de predicción del modelo de aprendizaje vs el de datos reales. Interpretar los resultados. |  |
| ¿? | ¿? | Preparación de datos previos a llevar a cabo un análisis. | Explicar y demostrar la funcionalidad con la función *dply*. | Demostrar cómo utilizar el comando *mutate*. |  |
|  |  |  |  | Demostrar una gráfica con una curva logística con *ggplot*. Explicar resultados. |  |
|  |  | Manejo del interfaz de RStudio | Demostrar cómo controlar la información que muestran los *chunks* (warnings, etc.) |  |  |
|  |  |  |  |  |  |
|  | Módulo de ayuda (no tiene definido una unidad todavía) | 1. Buscar ayuda para pruebas estadísticas de R.  2. Buscar ayuda para resolver errores o resolver mensajes de errores de pruebas estadísticas en R. | Localizar información sobre cómo hacer pruebas o métodos estadísticos utilizando un buscador de la Internet (DuckDuckGo, Google, etc.) | Demostrar de forma general. | Pendiente a ubicar en alguna unidad. |
|  |  | - Regresión lineal  - Pearson  - Spearman  - Kendall  - Correlacción  - Regresión logística  - t -test (Pareados y no pareados)  - ANOVA  - Kruskall Wallis  - Wilcoxon  - Mann-Whitney  - Tabla de contingencia |  |  |  |