

Segundo Parcial (30pts)

Programacion Estadística I

Lic. Alvaro Chirino Gutierrez

15/7/2020

Instrucciones

- Duración: 2 horas. 16:00 a 18:00
- Forma de entrega, incluir solo las respuestas en archivos separados, el nombre del archivo debe ser: respuesta_1, respuesta_2, etc.
- La entrega al correo achirino.stat@gmail.com máximo hasta las 18:00, pasado el tiempo no se tomara en cuenta el envío.
- Incluir en el correo su nombre completo.

Pregunta 1 (5 pts)

- (1pts) Describa la diferencia entre los procesos de inferencia: descriptiva, causal y predictivo
- (1pts) De un ejemplo de record linkage
- (2pts) Describa los errores de fila, columna y celda
- (1pts) Describa los errores asociados a las 3 V en big data

Pregunta 2 (10 pts)

- (5pts) Usando la base de datos de computo para las elecciones del 20 de octubre de 2019:
 1. Cargue la base en R
 2. Guarde la base en csv
 3. Cargue el csv como un objeto ffd. Con el nombre bd1.
 4. Sobre el objeto bd1 obtenga una tabla de departamento vs. tipo de elección
 5. Repita el paso 4 con una base obtenida con el comando as.ffdf o as.data.frame.ffdf
- (5pts) Usando la base de datos de computo para las elecciones del 20 de octubre de 2019. Seleccione solo Bolivia y las elecciones de presidente y vicepresidente.
 - Compare el tiempo de ejecución entre el código nativo de R y usando la librería parallel. Para obtener la suma de votos de las variables: “Inscritos, CC, FPV, MTS, UCS, MAS - IPSP, 21F, PDC, MNR, PAN-BOL, Votos Válidos, Blancos, Nulos”

Ejemplo de salida

##	Inscritos	CC	FPV	MTS	UCS
##	6973768	2186326	23028	75529	24007
##	MAS - IPSP	21F	PDC	MNR	PAN-BOL
##	2768427	256833	522576	41033	37841
##	Votos Válidos	Blancos	Nulos		
##	5935944	91610	222856		

Pregunta 3 (15 pts)

- (5 pts) Usando la encuesta de hogares defina una base de datos que contenga las siguientes variables para los jefes de hogar:
 - Edad
 - Sexo
 - Ingreso laboral
 - Departamento
 - Años de educación
 - Área
 - Incidencia de pobreza moderada (p0)
 - Acceso a internet en el hogar
 - Vivienda propia y totalmente pagada
 - Número de miembros en el hogar

(Sugerencia: Use la variable folio y el comando merge para unir bases)

- (10 pts) Para la base alojada en R y la base alojada en Spark, genere lo siguiente:
 - Promedio de años de educación del jefe del hogar por departamento y Área
 - Proporción de jefes del hogar con un ingreso laboral superior a 4000 Bs. Por Departamento y Sexo
 - Proporción de pobreza moderada en hogares con jefes de hogar de 30 años o menos por sexo
 - Promedio de miembros en el hogar por departamento, área y sexo del jefe del hogar
 - Gráfico sobre el acceso al internet por departamento (ggplot)