



Universidad Simón Bolívar
Departamento de Cómputo Científico y Estadística
Enero - Marzo 2019 (#Seguro)
Estadística para Ingenieros (CO3321)
Estadística para Matemáticos (CO3322)

Proyecto Final

Datos

Notas

Los datos adjuntos con el nombre **Proyecto2A.txt** o **Proyecto2B.txt**, contienen 27 observaciones de 9 variables, dichas observaciones o unidades muestrales corresponden a 28 estudiantes de cierto doctorado en España, y las variables a las características físicas de cada individuo. Las variables son:

- Estatura (centímetros, cm) (*estatura*)
- Peso (kilogramos, kg) (*peso*)
- Longitud de pie (cm) (*pie*)
- Longitud del brazo (cm) (*lbrazo*)
- Anchura de espalda (cm) (*anchoes*)
- Diámetro del cráneo (cm) (*dcraneo*)
- Longitud entre la rodilla y el tobillo (cm) (*lrodtob*)
- País de procedencia (*País_procedencia*).

Trabajo asignado

Suponiendo que la variable *estatura* es la variable dependiente:

1. Realice un análisis descriptivo y exploratorio de los datos. Incluya en este análisis la matriz de correlación.
2. Encuentre el modelo de regresión simple que mejor se ajuste a los datos; realice las pruebas estadísticas que considere conveniente para justificar su respuesta, incluyendo un análisis de residuales.

3. Emplee comparación de modelos para establecer el **modelo múltiple** más apropiado. Realice, como en el inciso 2, todas las pruebas estadísticas que considere conveniente para justificar su respuesta, incluyendo un análisis de residuos. Considere un nivel del 5%.
4. Estudios previos indican que los estudiantes de doctorado en España muestran un peso promedio de 64 kg, aunque estudios en otros doctorados suponen que dicho peso es superior al mostrado por este análisis. Con un nivel de confianza que usted considere necesario, realice un código en el software estadístico **R** que muestre el resultado de dicho análisis. Analice los resultados y concluya.
5. Para el modelo de regresión lineal simple obtenido en el inciso 2, realice la predicción correspondiente para la estatura de 5 estudiantes que se anexan a la muestra (Est=estudiantes), los datos se presentan en el Cuadro 1. Grafique los intervalos de predicción y de confianza respectivamente. Realice el análisis respectivo.
6. Existe suficiente evidencia que permita concluir que la estatura media de los estudiantes difiere con respecto al país de procedencia? Use el procedimiento de análisis de varianza para un diseño de un factor. Qué concluiría usted con un nivel de significancia $\alpha = 0.03$?
Sugerencia: Incluya un boxplot en el análisis.

NOTA: Realice transformaciones en las variables si lo considera necesario.

Variable	Est 1	Est 2	Est 3	Est 4	Est 5
Estatura					
Peso	43	55	70	90	65
Longitud de pie	34	37	39	44	36
Longitud de brazo	66	69	77	82	71
Anchura de espalda	35	40	49	52	43
Diámetro del cráneo	53	55	59	60	58
Longitud de rodilla y tobillo	37	40	50	52	49

Table 1: Datos que se anexan a los datos originales

Asignación de Grupos

Proyecto2A

Grupo 1		Grupo 2		Grupo 3	
15-10463	Manuel Faria	16-10072	Amin Arriaga	14-10272	Cristian Da Silva
15-11041	Juan Oropeza	16-10400	Ángel Garces	14-10628	Leonardo Martin
13-11341	David Segura	15-10639	Wilfredo Graterol	10-10705	Edwin Sosa
Grupo 4		Grupo 5		Grupo 6	
15-10460	Pedro Fagundez	12-11499	Orlando Chaparro	13-10406	Gabriel Durán
15-11523	Neil Villamizar	12-11163	Francisco Márquez	11-10303	Nilson Estrada
15-11551	Arturo Yépez	11-10629	Jesús Molina	11-11073	Laura Villalba
Grupo 7		Grupo 8		Grupo 9	
15-10264	Mariela Castro	15-10034	Manuel Alvarado	15-10360	Jesús De Aguiar
15-11099	Josman Peralta	14-10459	Fabian Guiliani	15-10732	Nicolás Jaua
15-11440	Samuel Tovar	15-10800	Guillermo López	15-10778	Marcos Lerones
Grupo 10		Grupo 11		Grupo 12	
15-10686	Ricardo Hernández	14-10210	Allison Centeno	15-10172	Christopher Bolívar
15-11239	Ricardo Rodríguez	12-10400	Javier Medina	15-10718	Pietro Iaia
15-11470	Manuel De Sousa	15-11201	Andrea Reyes	15-11579	Daniela Zorrilla
		Grupo 13			
		14-10363	Johan Franco		
		15-10540	Luis García		
		15-11574	Abel Zavaleta		

Proyecto2B

Grupo 1		Grupo 2		Grupo 3	
14-10032	Luis Alvarado	13-10697	Reinaldo Jiménez	14-10169	Salvador Canzoneri
14-10193	Manuel Castillo	11-11521	Wilfredo Torres	14-10283	Oleamny De Almada
14-10544	Andrés Kowalski	13-11494	Adriana Villegas	15-10749	Fabiana Kilzi
Grupo 4		Grupo 5		Grupo 6	
14-10101	Roberto Bernardi	15-10276	David Cedeño	11-10390	Manuel González
15-10898	Andreina Marval	15-10713	Ruby Hinojosa	13-10640	Victor Hernández
14-10951	María Rodríguez	15-11101	Maria Perdigon	12-11337	Daniela Torres
Grupo 7		Grupo 8		Grupo 9	
12-11409	Gianni D'Marco	14-10237	Miguel Colorado	15-10123	José Barrera
11-10365	Rosana García	10-10548	Pedro Pérez	15-11095	Diego Peña
13-10931	Ángel Morante	13-11524	Roxanna Zapata	15-11550	Jean Yazbek
Grupo 10		Grupo 11		Grupo 12	
15-10611	Carlos González	14-10380	Juan García	15-10332	Jesyreth Corredor
15-10696	María Hernández	12-11119	Carlos Leal	15-10663	Shannon Gutt
15-11196	Antonella Requena	14-11108	Sergio Valero	15-11296	Valentina Rosas

Criterios de corrección para el proyecto

La estructura que debe tener el informe es:

- Portada con resumen (en la misma hoja).
- Planteamiento del problema (incluyendo los objetivos del trabajo), descripción de la base de datos y la metodología a emplear.
- Desarrollo (donde se realizan las asignaciones).
- Conclusiones y recomendaciones.
- Bibliografía.
- Anexos (+ códigos en R).

En la portada se debe encontrar el título del proyecto, el resumen y la identificación de los autores.

Una de las partes más importantes del informe es el resumen; en este se deben plantear los objetivos del proyecto y una breve descripción de la base de datos y de la metodología empleada. También se deben encontrar los resultados del proyecto (o por lo menos, los más substanciales), y se debe aclarar las implicaciones de estos resultados, las conclusiones y recomendaciones (simplificadas) que hace el analista.

El cuerpo principal del informe, debe comenzar con el planteamiento del problema, y luego describir la base de datos y la metodología que se empleará durante la resolución del mismo. Se deben usar tablas y gráficos para facilitar la lectura del informe y obtener la atención del cliente; las tablas y gráficos deben estar comentadas, no se permiten tablas o gráficos a las que no se hacen referencia. Debido a que el informe no debe tener más de diez (10) páginas (desde la portada a la bibliografía), se debe resumir la información en tablas o diagramas y se deben seleccionar los gráficos más relevantes.

En las conclusiones se presentan los resultados obtenidos conjuntamente con las implicaciones que tienen esos resultados (sin profundizar en terrenos del área en el que se desenvuelve el cliente, a menos de que se esté seguro del impacto de las implicaciones). Recuerde que este es un trabajo parecido al de asesoría y que el cliente es el que toma las decisiones, el analista sólo plantea alternativas y puede sugerir alguna de las soluciones al problema.

Presentación de resultados

- Presente sus resultados en tablas ordenadas e interprete.
- Identifique en los diagramas de caja si hay datos atípicos, cómo es la distribución de los datos, si es sesgada a la derecha, etc.
- Los gráficos tiene que tener su título y los nombres de los ejes (todo en español).

Es INACEPTABLE

- No se aceptará presentación de resultados con manuscritos escaneados.
- Se anulará la evaluación de aquellos que compartan fotografías tomadas desde la pantalla de la computadora.
- No se aceptará un copy y paste de los resultados.
- No se aceptarán títulos de las gráficas generados por defecto en el programa.

Ejemplos

Por último se exponen unos ejemplos para la presentación de los resultados (Gráficas y Tablas), para mayor información se puede consultar las normas de la Universidad Simón Bolívar para la elaboración de trabajos.

Tabla 1. Resumen estadístico para la variable Índice de Aprovechamiento

Resumen Estadístico							
Variable	Mínimo	Primer	Mediana	Media	Tercer	Máximo	Desviación
IAP	0	1	2.14	1.97	2.95	4.05	1.16

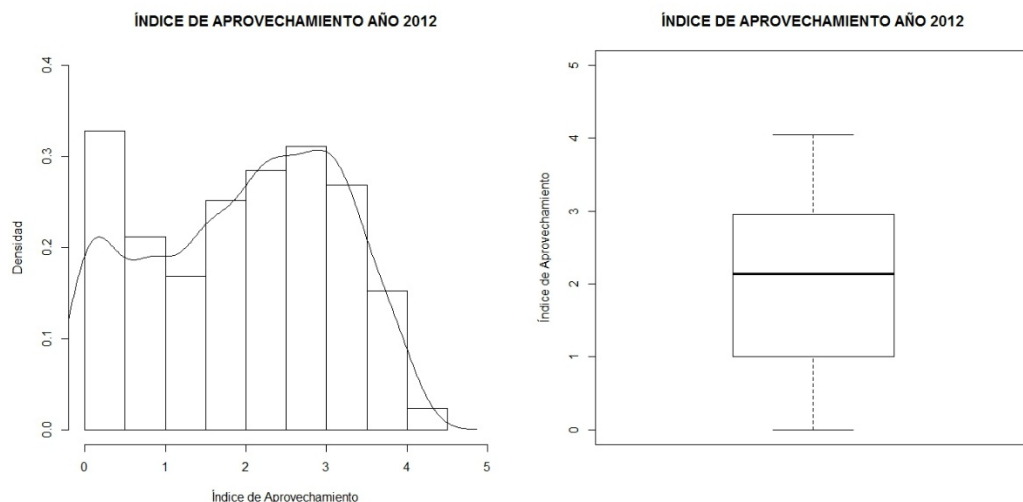


Gráfico 1. Histograma y gráfico de caja para la variable Índice de aprovechamiento.

NOTA: recuerde que existen normas para la elaboración de trabajos propias de la USB, es recomendable revisar las mismas para la escritura del proyecto. Por ejemplo, es muy común cometer errores en la bibliografía. Recuerde que el autor debe ser mencionado en el texto, y posteriormente señalar la referencia en la bibliografía.

Ejemplo:

Para Gelman y otros (2014), el muestreador de Gibbs es un método de gran utilidad en problemas donde el espacio de parámetros es multidimensional...

En este trabajo se aplicó el programa R Development Core Team (2015)...

Según Gil, J. (s/f), los métodos...

En la bibliografía

Gelman, A., Carlin, J., Stern, H. y Rubin, D. (2004). Bayesian data analysis. Second Edition. Chapman & Hall/ CRC.

Gil, J. (s/f). Modelos de medición: desarrollos actuales, supuestos, ventajas e inconvenientes. Universidad de Sevilla. [Revista en Línea]. Disponible: <http://innoevalua.us.es/files/irt.pdf> [Consulta: 2015, Diciembre, 09].

R Development Core Team (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, Disponible: <http://www.R-project.org>.

Condiciones de entrega

- a El informe debe ser entregado en forma electrónica y en formato “.pdf”.
- b La entrega se realizará al correo electrónico povallesgarcia@usb.ve a más tardar el viernes 31 de mayo de 2019 a las 10:00 a.m. El asunto del correo DEBE ser: “*Proyecto. CO3321*”.
- c No se corregirán informes entregados fuera del tiempo establecido para la entrega.