



Universidad Simón Bolívar
Departamento de Cómputo Científico y Estadística
Trimestre Abril-Julio 2018
Estadística para Ingenieros-CO3321

Proyecto Final

Datos

Estudiantes

Los datos *estudiantes.txt*, contiene 300 observaciones de 9 variables. Las observaciones o unidades muestrales corresponden a estudiantes venezolanos que realizan pruebas estandarizadas. Los datos que se registran para cada estudiante son:

“Mat” = Nota en la prueba de matemáticas.
“Soc” = Nota en la prueba de ciencias sociales.
“Len” = Nota en la prueba de lenguaje y literatura.
“Cult” = Nota en la prueba de cultura general.
“Cien” = Nota en la prueba de ciencias.
“Log” = Nota en la prueba de razonamiento lógico.
“Esp” = Nota en la prueba de habilidad espacial.
“Prom” = Promedio de bachillerato.
Zona del aspirante

Estudio de aluminio

El archivo de datos *proceso.txt* contiene 220 muestras recogidas en ciertos centros de investigación en los cuales se estudia la formación de óxido de aluminio. Cada observación contiene información sobre 9 variables. Estas son:

“y” = Porcentaje de óxido de aluminio que precipita.
“x1” = Concentración del ácido.
“x2” = Valor final de pH de la solución en unidades de pH.
“x3” = Temperatura del proceso en grados centígrados.
“x4” = Concentración de la base.
“x5” = Velocidad de agitación en rpm.
“x6” = Velocidad de adición de la base en mililitros por hora.
“x7” = Tiempo del proceso en horas.
“Cent” = Centro de investigación donde se toma la muestra.

Trabajo asignado

Estudiantes

Para este grupo de variables se solicita el siguiente trabajo:

1. Realizar un análisis descriptivo de los datos.
2. Realizar un gráfico de dispersión y una matriz de correlación de las variables.
3. Realizar una prueba de bondad de ajuste para determinar si la variable “prom” se distribuye en forma normal.
4. Realice un intervalo de confianza del 97% para cada variable en estudio.
5. Halle un modelo lineal que explique mejor la variable “prom”. Incluya todas las pruebas necesarias para llegar a este modelo, así como un análisis de residuos del modelo final.
6. Con los datos *estudiantes_pre.txt* haga una predicción de la variable “prom” (con el mejor modelo) y haga un histograma, diagrama de cajas y resumen estadístico de los residuos de predicción (valor observado vs. predicción del modelo) para concluir con relación al poder predictivo del modelo.
7. Realice un análisis de varianza para decidir si las medias por zona son iguales, para cada variable.
8. En el caso de obtener en el análisis de varianza que existe una diferencia significativa, por medio de pruebas de hipótesis, decidir cuales son las zonas con diferencias.

Estudio de aluminio

Para este grupo de variables se solicita el siguiente trabajo:

1. Realizar un análisis descriptivo de los datos.
2. Realizar un gráfico de dispersión y una matriz de correlación de las variables.
3. Realizar una prueba de bondad de ajuste para determinar si la variable “y” se distribuye en forma normal.
4. Realice un intervalo de confianza del 93% para cada variable en estudio.
5. Halle un modelo lineal que explique mejor la variable “y”. Incluya todas las pruebas necesarias para llegar a este modelo, así como un análisis de residuos del modelo final.
6. Con los datos *proceso_pre.txt* haga una predicción de la variable “y” (con el mejor modelo) y haga un histograma, diagrama de cajas y resumen estadístico de los residuos de predicción (valor observado vs. predicción del modelo) para concluir con relación al poder predictivo del modelo.
7. Realice un análisis de varianza para decidir si las medias por centro son iguales, para cada variable.
8. En el caso de obtener en el análisis de varianza que existe una diferencia significativa, por medio de pruebas de hipótesis, decidir cuales son las zonas con diferencias.

Asignación de Grupos

Estudiantes

Grupo 1		Grupo 2		Grupo 3	
14-10613	Lorena Marin	12-11170	Andrés Cardenas	11-10278	Albert Díaz
13-11295	María Salas	12-10863	Daniel Francis	12-11573	Diego Fermin
13-11303	Abelardo Salazar	11-11431	Aimee Salazar	12-11468	Victoria Torres
Grupo 4		Grupo 5		Grupo 6	
12-11499	Orlando Chaparro	11-10106	Emilio Blanco	14-10261	Alfredo Cruz
05-38194	Ivanhoe Gamarra	10-10488	Edwin Murillo	14-10339	Manuel Fernández
09-10870	Edgar Valderrama				
		Grupo 7			
		13-10173	José Bracuto		
		14-10215	Andrés Chacon		
		13-11156	Rafael Cisneros		

Proceso

Grupo 1		Grupo 2		Grupo 3	
12-10979	Sebastián Clavell	12-11409	Gianni D'Marco	12-11285	Raúl Bander
12-11555	Erick Flejan	14-10672	Andrés Mendoza	14-10102	Bárbara Berroterán
13-11089	Carlos Pérez	12-11337	Daniela Torres	14-10392	Alimi Garmendia
Grupo 4		Grupo 5		Grupo 6	
14-10193	Manuel Castillo	09-10810	Paul Baptista	13-10268	Kevin Chacón
14-10459	Fabián Guiliani	14-10755	María Oliveros	12-10679	Grey Rondón
13-10943	Verónica Moreno	13-11104	Gabriel Pérez	10-10705	Edwin Sosa
		Grupo 7			
		09-10066	Gabriel Austin		
		12-10921	Nathalia Silvera		
		14-11127	Miguel Vélez		

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

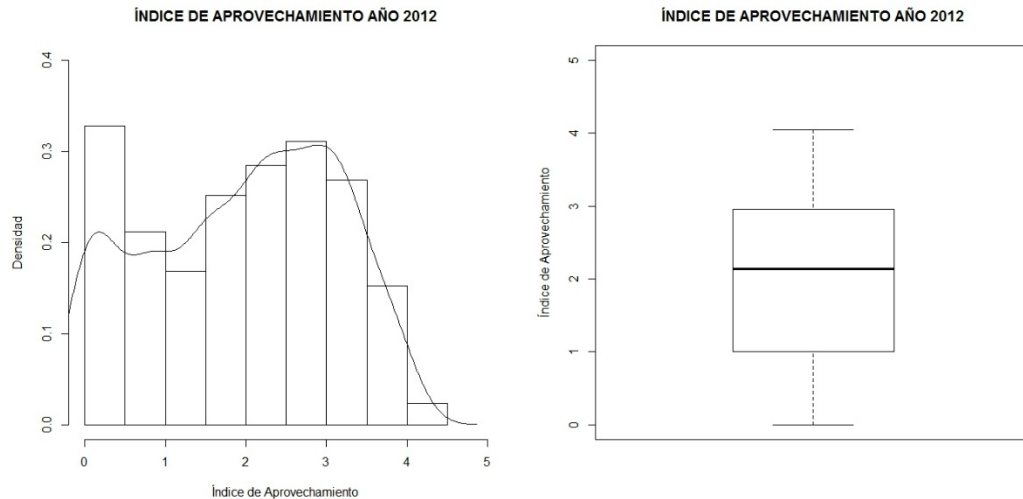


Gráfico 1. Histograma y gráfico de caja para la variable índice de aprovechamiento.

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 6 de julio de 2018. El asunto del correo DEBE ser: “*Proyecto. CO3321*”.
- c No se corregirá informes entregados fuera del tiempo establecido para la entrega.